

---

# PRINCIPLES OF COMMUNICATION SYSTEMS

---

Second Edition

**Herbert Taub**

**Donald L. Schilling**

*Professors of Electrical Engineering  
The City College of New York*

**McGraw-Hill Publishing Company**

New York St. Louis San Francisco Auckland Bogotá  
Caracas Hamburg Lisbon London Madrid Mexico Milan  
Montreal New Delhi Oklahoma City Paris San Juan  
São Paulo Singapore Sydney Tokyo Toronto

## SPECTRAL ANALYSIS

### INTRODUCTION

Suppose that two people, separated by a considerable distance, wish to communicate with one another. If there is a pair of conducting wires extending from one location to another, and if each place is equipped with a microphone and ear-piece, the communication problem may be solved. The microphone, at one end of the wire communications channel, impresses an electric signal voltage on the line, which voltage is then received at the other end. The received signal, however, will have associated with it an erratic, random, unpredictable voltage waveform which is described by the term *noise*. The origin of this noise will be discussed more fully in Chaps. 7 and 14. Here we need but to note that at the atomic level the universe is in a constant state of agitation, and that this agitation is the source of a very great deal of this noise. Because of the length of the wire link, the received message signal voltage will be greatly attenuated in comparison with its level at the transmitting end of the link. As a result, the message signal voltage may not be very large in comparison with the noise voltage, and the message will be perceived with difficulty or possibly not at all. An amplifier at the receiving end will not solve the problem, since the amplifier will amplify signal and noise alike. As a matter of fact, as we shall see, the amplifier itself may well be a source of additional noise.

A principal concern of communication theory and a matter which we discuss extensively in this book is precisely the study of methods to suppress, as far as possible, the effect of noise. We shall see that, for this purpose, it may be better not to transmit directly the original signal (the microphone output in our

example). Instead, the original signal is used to generate a different signal waveform, which new signal waveform is then impressed on the line. This processing of the original signal to generate the transmitted signal is called *encoding* or *modulation*. At the receiving end an inverse process called *decoding* or *demodulation* is required to recover the original signal.

It may well be that there is a considerable expense in providing the wire communication link. We are, therefore, naturally led to inquire whether we may use the link more effectively by arranging for the simultaneous transmission over the link of more than just a single waveform. It turns out that such multiple transmission is indeed possible and may be accomplished in a number of ways. Such simultaneous multiple transmission is called *multiplexing* and is again a principal area of concern of communication theory and of this book. It is to be noted that when wire communications links are employed, then, at least in principle, separate links may be used for individual messages. When, however, the communications medium is free space, as in radio communication from antenna to antenna, multiplexing is essential.

In summary then, *communication theory* addresses itself to the following questions: Given a communication channel, how do we arrange to transmit as many simultaneous signals as possible, and how do we devise to suppress the effect of noise to the maximum extent possible? In this book, after a few mathematical preliminaries, we shall address ourselves precisely to these questions, first to the matter of multiplexing, and thereafter to the discussion of noise in communications systems.

A branch of mathematics which is of inestimable value in the study of communications systems is *spectral analysis*. Spectral analysis concerns itself with the description of waveforms in the *frequency domain* and with the correspondence between the frequency-domain description and the time-domain description. It is assumed that the reader has some familiarity with spectral analysis. The presentation in this chapter is intended as a review, and will serve further to allow a compilation of results which we shall have occasion to use throughout the remainder of this text.

## 1.1 FOURIER SERIES<sup>1</sup>

A periodic function of time  $v(t)$  having a fundamental period  $T_0$  can be represented as an infinite sum of sinusoidal waveforms. This summation, called a *Fourier series*, may be written in several forms. One such form is the following:

$$v(t) = A_0 + \sum_{n=1}^{\infty} A_n \cos \frac{2\pi n t}{T_0} + \sum_{n=1}^{\infty} B_n \sin \frac{2\pi n t}{T_0} \quad (1.1-1)$$

The constant  $A_0$  is the average value of  $v(t)$  given by

$$A_0 = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} v(t) dt \quad (1.1-2)$$

while the coefficients  $A_n$  and  $B_n$  are given by

$$A_n = \frac{2}{T_0} \int_{-T_0/2}^{T_0/2} v(t) \cos \frac{2\pi n t}{T_0} dt \quad (1.1-3)$$

and

$$B_n = \frac{2}{T_0} \int_{-T_0/2}^{T_0/2} v(t) \sin \frac{2\pi n t}{T_0} dt \quad (1.1-4)$$

An alternative form for the Fourier series is

$$v(t) = C_0 + \sum_{n=1}^{\infty} C_n \cos \left( \frac{2\pi n t}{T_0} - \phi_n \right) \quad (1.1-5)$$

where  $C_0$ ,  $C_n$ , and  $\phi_n$  are related to  $A_0$ ,  $A_n$ , and  $B_n$  by the equations

$$C_0 = A_0 \quad (1.1-6a)$$

$$C_n = \sqrt{A_n^2 + B_n^2} \quad (1.1-6b)$$

and

$$\phi_n = \tan^{-1} \frac{B_n}{A_n} \quad (1.1-6c)$$

The Fourier series of a periodic function is thus seen to consist of a summation of harmonics of a fundamental frequency  $f_0 = 1/T_0$ . The coefficients  $C_n$  are called *spectral amplitudes*; that is,  $C_n$  is the amplitude of the *spectral component*  $C_n \cos(2\pi n f_0 t - \phi_n)$  at frequency  $n f_0$ . A typical *amplitude spectrum* of a periodic waveform is shown in Fig. 1.1-1a. Here, at each harmonic frequency, a vertical line has been drawn having a length equal to the spectral amplitude associated with each harmonic frequency. Of course, such an amplitude spectrum, lacking the phase information, does not specify the waveform  $v(t)$ .

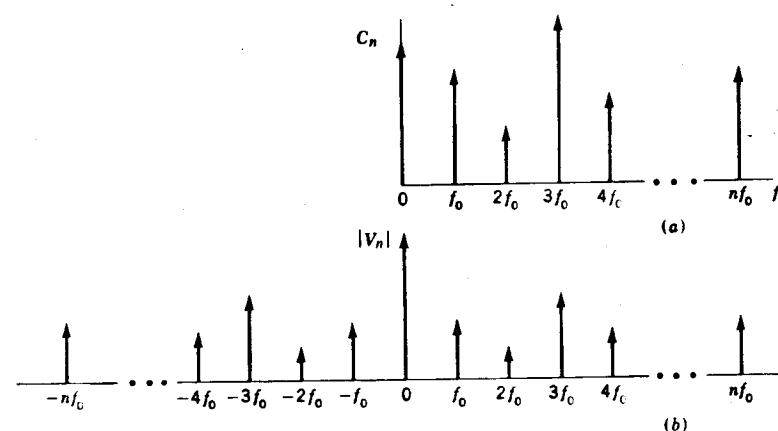


Figure 1.1-1 (a) A one-sided plot of spectral amplitude of a periodic waveform. (b) The corresponding two-sided plot.

## 1.2 EXPONENTIAL FORM OF THE FOURIER SERIES

The exponential form of the Fourier series finds extensive application in communication theory. This form is given by

$$v(t) = \sum_{n=-\infty}^{\infty} V_n e^{j2\pi n t / T_0} \quad (1.2-1)$$

where  $V_n$  is given by

$$V_n = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} v(t) e^{-j2\pi n t / T_0} dt \quad (1.2-2)$$

The coefficients  $V_n$  have the property that  $V_n$  and  $V_{-n}$  are complex conjugates of one another, that is,  $V_n = V_{-n}^*$ . These coefficients are related to the  $C_n$ 's in Eq. (1.1-5) by

$$V_0 = C_0 \quad (1.2-3a)$$

$$V_n = \frac{C_n}{2} e^{-j\phi_n} \quad (1.2-3b)$$

The  $V_n$ 's are the *spectral amplitudes of the spectral components*  $V_n e^{j2\pi n f_0 t}$ . The amplitude spectrum of the  $V_n$ 's shown in Fig. 1.1-1b corresponds to the amplitude spectrum of the  $C_n$ 's shown in Fig. 1.1-1a. Observe that while  $V_0 = C_0$ , otherwise each spectral line in 1.1-1a at frequency  $f$  is replaced by the 2 spectral lines in 1.1-1b, each of half amplitude, one at frequency  $f$  and one at frequency  $-f$ . The amplitude spectrum in 1.1-1a is called a *single-sided* spectrum, while the spectrum in 1.1-1b is called a *two-sided* spectrum. We shall find it more convenient to use the two-sided amplitude spectrum and shall consistently do so from this point on.

## 1.3 EXAMPLES OF FOURIER SERIES

A waveform in which we shall have occasion to have some special interest is shown in Fig. 1.3-1a. The waveform consists of a periodic sequence of impulses of strength  $I$ . As a matter of convenience we have selected the time scale so that an impulse occurs at  $t = 0$ . The impulse at  $t = 0$  is written as  $I \delta(t)$ . Here  $\delta(t)$  is the delta function which has the property that  $\delta(t) = 0$  except when  $t = 0$  and further

$$\int_{-\infty}^{\infty} \delta(t) dt = 1 \quad (1.3-1)$$

The strength of an impulse is equal to the area under the impulse. Thus the strength of  $\delta(t)$  is 1, and the strength of  $I \delta(t)$  is  $I$ .

The periodic impulse train is written

$$v(t) = I \sum_{k=-\infty}^{\infty} \delta(t - kT_0) \quad (1.3-2)$$

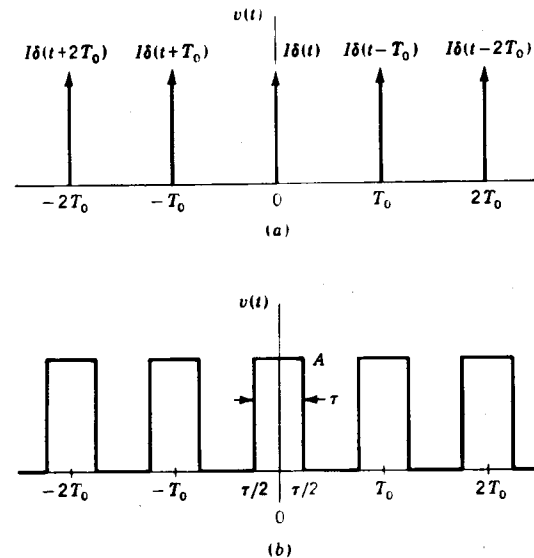


Figure 1.3-1 Examples of periodic functions. (a) A periodic train of impulses. (b) A periodic train of pulses of duration  $\tau$ .

We then find, using Eqs. (1.1-2) and (1.1-3),

$$A_0 = \frac{I}{T_0} \int_{-T_0/2}^{T_0/2} \delta(t) dt = \frac{I}{T_0} \quad (1.3-3)$$

$$A_n = \frac{2I}{T_0} \int_{-T_0/2}^{T_0/2} \delta(t) \cos \frac{2\pi n t}{T_0} dt = \frac{2I}{T_0} \quad (1.3-4)$$

and, using Eq. (1.1-4),

$$B_n = \frac{2I}{T_0} \int_{-T_0/2}^{T_0/2} \delta(t) \sin \frac{2\pi n t}{T_0} dt = 0 \quad (1.3-5)$$

Further we have, using Eq. (1.1-6),

$$C_0 = \frac{I}{T_0} \quad C_n = \frac{2I}{T_0} \quad \phi_n = 0 \quad (1.3-6)$$

and from Eq. (1.2-3)

$$V_0 = V_n = \frac{I}{T_0} \quad (1.3-7)$$

Hence  $v(t)$  may be written in the forms

$$\begin{aligned} v(t) &= I \sum_{k=-\infty}^{\infty} \delta(t - kT_0) = \frac{I}{T_0} + \frac{2I}{T_0} \sum_{n=1}^{\infty} \cos \frac{2\pi nt}{T_0} \\ &= \frac{I}{T_0} \sum_{n=-\infty}^{\infty} e^{j2\pi nt/T_0} \end{aligned} \quad (1.3-8)$$

As a second example, let us find the Fourier series for the periodic train of pulses of amplitude  $A$  and duration  $\tau$  as shown in Fig. 1.3-1b. We find

$$A_0 = C_0 = V_0 = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} v(t) dt = \frac{A\tau}{T_0} \quad (1.3-9)$$

$$\begin{aligned} A_n = C_n = 2V_n &= \frac{2}{T_0} \int_{-T_0/2}^{T_0/2} v(t) \cos \frac{2\pi nt}{T_0} dt \\ &= \frac{2A\tau}{T_0} \frac{\sin(n\pi\tau/T_0)}{n\pi\tau/T_0} \end{aligned} \quad (1.3-10)$$

and

$$B_n = 0 \quad \phi_n = 0 \quad (1.3-11)$$

Thus

$$v(t) = \frac{A\tau}{T_0} + \frac{2A\tau}{T_0} \sum_{n=1}^{\infty} \frac{\sin(n\pi\tau/T_0)}{n\pi\tau/T_0} \cos \frac{2\pi nt}{T_0} \quad (1.3-12a)$$

$$= \frac{A\tau}{T_0} \sum_{n=-\infty}^{\infty} \frac{\sin(n\pi\tau/T_0)}{n\pi\tau/T_0} e^{j2\pi nt/T_0} \quad (1.3-12b)$$

Suppose that in the waveform of Fig. 1.3-1b we reduce  $\tau$  while adjusting  $A$  so that  $A\tau$  is a constant, say  $A\tau = I$ . We would expect that in the limit, as  $\tau \rightarrow 0$ , the Fourier series for the pulse train in Eq. (1.3-12) should reduce to the series for the impulse train in Eq. (1.3-8). It is readily verified that such is indeed the case since as  $\tau \rightarrow 0$

$$\frac{\sin(n\pi\tau/T_0)}{n\pi\tau/T_0} \rightarrow 1 \quad (1.3-13)$$

## 1.4 THE SAMPLING FUNCTION

A function frequently encountered in spectral analysis is the sampling function  $Sa(x)$  defined by

$$Sa(x) \equiv \frac{\sin x}{x} \quad (1.4-1)$$

[A closely related function is  $\text{sinc } x$  defined by  $\text{sinc } x = (\sin \pi x)/\pi x$ .] The function  $Sa(x)$  is plotted in Fig. 1.4-1. It is symmetrical about  $x = 0$ , and at  $x = 0$  has

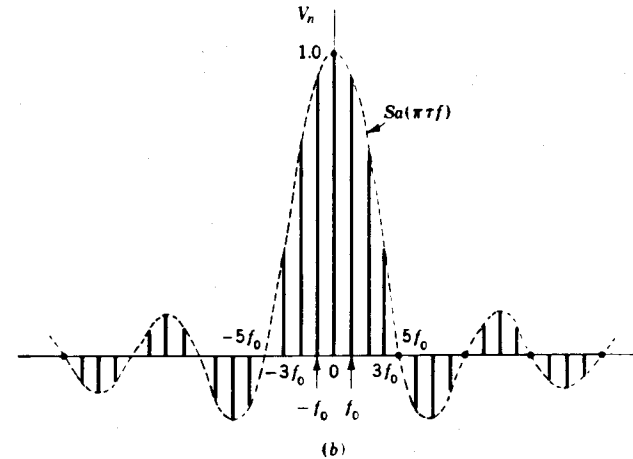
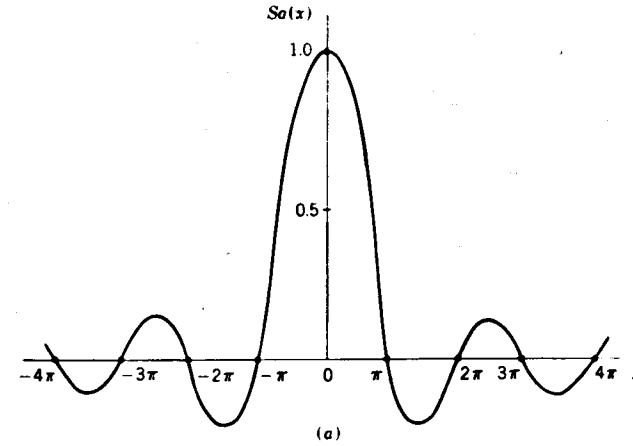


Figure 1.4-1 (a) The function  $Sa(x)$ . (b) The spectral amplitudes  $V_n$  of the two-sided Fourier representation of the pulse train of Fig. 1.3-1b for  $A = 4$  and  $\tau/T_0 = \frac{1}{4}$ .

the value  $Sa(0) = 1$ . It oscillates with an amplitude that decreases with increasing  $x$ . The function passes through zero at equally spaced intervals at values of  $x = \pm n\pi$ , where  $n$  is an integer other than zero. Aside from the peak at  $x = 0$ , the maxima and minima occur approximately midway between the zeros, i.e., at  $x = \pm(n + \frac{1}{2})\pi$ , where  $|\sin x| = 1$ . This approximation is poorest for the minima closest to  $x = 0$  but improves as  $x$  becomes larger. Correspondingly, the approximate value of  $Sa(x)$  at these extremal points is

$$Sa[\pm(n + \frac{1}{2})\pi] = \frac{2(-1)^n}{(2n + 1)\pi} \quad (1.4-2)$$

We encountered the sampling function in the preceding section in Eq. (1.3-12) which expresses the spectrum of a periodic sequence of rectangular pulses. In that equation we have  $x = n\pi\tau/T_0$ . The spectral amplitudes of Eq. (1.3-12b) are plotted in Fig. 1.4-1b for the case  $A = 4$  and  $\tau/T_0 = \frac{1}{4}$ . The spectral components appear at frequencies which are multiples of the fundamental frequency  $f_0 = 1/T_0$ , that is, at frequencies  $f = nf_0 = n/T_0$ . The envelope of the spectral components of  $Sa(\pi\tau f)$  is also shown in the figure. Here we have replaced  $x$  by

$$x = n\pi\tau/T_0 = \pi\tau nf_0 = \pi\tau f \quad (1.4-3)$$

## 1.5 RESPONSE OF A LINEAR SYSTEM

The Fourier trigonometric series is by no means the only way in which a periodic function may be expanded in terms of other functions.<sup>2</sup> As a matter of fact, the number of such possible alternative expansions is limitless. However, what makes the Fourier trigonometric expansion especially useful is the distinctive and unique characteristic of the sinusoidal waveform, this characteristic being that when a sinusoidal excitation is applied to a linear system, the response everywhere in the system is similarly sinusoidal and has the same frequency as the excitation. That is, the sinusoidal waveform preserves its waveshape. And since the waveshape is preserved, then, in order to characterize the relationship of the response to the excitation, we need but to specify how the response amplitude is related to the excitation amplitude and how the response phase is related to the excitation phase. Therefore with sinusoidal excitation, two numbers (amplitude ratio and phase difference) are all that are required to deduce a response. It turns out to be possible and very convenient to incorporate these two numbers into a single complex number.

Let the input to a linear system be the spectral component

$$v_i(t, \omega_n) = V_n e^{j2\pi nt/T_0} = V_n e^{j\omega_n t} \quad (1.5-1)$$

The waveform  $v_i(t, \omega_n)$  may be, say, the voltage applied to the input of an electrical filter as in Fig. 1.5-1. Then the filter output  $v_o(t, \omega_n)$  is related to the input by a complex transfer function

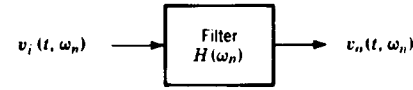
$$H(\omega_n) = |H(\omega_n)| e^{-j\theta(\omega_n)} \quad (1.5-2)$$

that is, the output is

$$\begin{aligned} v_o(t, \omega_n) &= H(\omega_n) v_i(t, \omega_n) = |H(\omega_n)| e^{-j\theta(\omega_n)} V_n e^{j\omega_n t} \\ &= |H(\omega_n)| V_n e^{j[\omega_n t - \theta(\omega_n)]} \end{aligned} \quad (1.5-3)$$

Actually, the spectral component in Eq. (1.5-1) is not a physical voltage. Rather, the physical input voltage  $v_{ip}(t)$  which gives rise to this spectral component is the sum of this spectral component and its complex conjugate, that is,

$$v_{ip}(t, \omega_n) = V_n e^{j\omega_n t} + V_n^* e^{-j\omega_n t} = V_n e^{j\omega_n t} + V_n^* e^{-j\omega_n t} = 2\text{Re}(V_n e^{j\omega_n t}) \quad (1.5-4)$$



**Figure 1.5-1** A sinusoidal waveform  $v_i(t, \omega_n)$  of angular frequency  $\omega_n$  is applied at the input to a network (filter) whose transfer characteristic at  $\omega_n$  is  $H(\omega_n)$ . The output  $v_o(t, \omega_n)$  differs from the input only in amplitude and phase.

The corresponding physical output voltage is  $v_{op}(t, \omega_n)$  given by

$$v_{op}(t, \omega_n) = H(\omega_n) V_n e^{j\omega_n t} + H(-\omega_n) V_n^* e^{-j\omega_n t} \quad (1.5-5)$$

Since  $v_{op}(t, \omega_n)$  must be real, the two terms in Eq. (1.5-5) must be complex conjugates, and hence we must have that  $H(\omega_n) = H^*(-\omega_n)$ . Therefore, since  $H(\omega_n) = |H(\omega_n)| e^{-j\theta(\omega_n)}$ , we must have that

$$|H(\omega_n)| = |H(-\omega_n)| \quad (1.5-6)$$

and

$$\theta(\omega_n) = -\theta(-\omega_n) \quad (1.5-7)$$

that is,  $|H(\omega_n)|$  must be an even function and  $\theta(\omega_n)$  an odd function of  $\omega_n$ .

If, then, an excitation is expressed as a Fourier series in exponential form as in Eq. (1.2-1), the response is

$$v_o(t) = \sum_{n=-\infty}^{\infty} H(\omega_n) V_n e^{j2\pi nt/T_0} \quad (1.5-8)$$

If the form of Eq. (1.1-5) is used, the response is

$$v_o(t) = H(0)C_0 + \sum_{n=1}^{\infty} |H(\omega_n)| C_n \cos \left[ \frac{2\pi nt}{T_0} - \phi_n - \theta(\omega_n) \right] \quad (1.5-9)$$

Given a periodic waveform, the coefficients in the Fourier series may be evaluated. Thereafter, if the transfer function  $H(\omega_n)$  of a system is known, the response may be written out formally as in, say, Eq. (1.5-8) or Eq. (1.5-9). Actually these equations are generally of small value from a computational point of view. For, except in rather special and infrequent cases, we should be hard-pressed indeed to recognize the waveform of a response which is expressed as the sum of an infinite (or even a large) number of sinusoidal terms. On the other hand, the concept that a response may be written in the form of a linear superposition of responses to individual spectral components, as, say, in Eq. (1.5-8), is of inestimable value.

## 1.6 NORMALIZED POWER

In the analysis of communication systems, we shall often find that, given a waveform  $v(t)$ , we shall be interested in the quantity  $\bar{v^2(t)}$  where the bar indicates the time-average value. In the case of periodic waveforms, the time averaging is done over one cycle. If, in our mind's eye, we were to imagine that the waveform  $v(t)$  appear across a 1-ohm resistor, then the power dissipated in that resistor would

be  $\overline{v^2(t)}$  volts<sup>2</sup>/1 ohm =  $W$  watts, where the number  $W$  would be numerically equal to the numerical value of  $\overline{v^2(t)}$ , the mean-square value of  $v(t)$ . For this reason  $\overline{v^2(t)}$  is generally referred to as the *normalized power* of  $v(t)$ . It is to be kept in mind, however, that the dimension of normalized power is volts<sup>2</sup> and not watts. When, however, no confusion results from so doing, we shall often follow the generally accepted practice of dropping the word "normalized" and refer instead simply to "power." We shall often have occasion also to calculate *ratios* of normalized powers. In such cases, even if the dimension "watts" is applied to normalized power, no harm will have been done, for the dimensional error in the numerator and the denominator of the ratio will cancel out.

Suppose that in some system we encounter at one point or another normalized powers  $S_1$  and  $S_2$ . If the ratio of these powers is of interest, we need but to evaluate, say,  $S_2/S_1$ . It frequently turns out to be more convenient not to specify this ratio directly but instead to specify the quantity  $K$  defined by

$$K \equiv 10 \log \frac{S_2}{S_1} \quad (1.6-1)$$

Like the ratio  $S_2/S_1$ , the quantity  $K$  is dimensionless. However, in order that one may know whether, in specifying a ratio we are stating the number  $S_2/S_1$  or the number  $K$ , the term *decibel* (abbreviated dB) is attached to the number  $K$ . Thus, for example, suppose  $S_2/S_1 = 100$ , then  $\log S_2/S_1 = 2$  and  $K = 20$  dB. The advantages of the use of the decibel are twofold. First, a very large power ratio may be expressed in decibels by a much smaller and therefore often more convenient number. Second, if power ratios are to be multiplied, such multiplication may be accomplished by the simpler arithmetic operation of addition if the ratios are first expressed in decibels. Suppose that  $S_2$  and  $S_1$  are, respectively, the normalized power associated with sinusoidal signals of amplitudes  $V_2$  and  $V_1$ . Then  $S_2 = V_2^2/2$ ,  $S_1 = V_1^2/2$ , and

$$K = 10 \log \frac{V_2^2/2}{V_1^2/2} = 20 \log \frac{V_2}{V_1} \quad (1.6-2)$$

The use of the decibel was introduced in the early days of communications systems in connection with the transmission of signals over telephone lines. (The "bel" in decibel comes from the name Alexander Graham Bell.) In those days the decibel was used for the purpose of specifying ratios of *real* powers, not normalized powers. Because of this early history, occasionally some confusion occurs in the meaning of a ratio expressed in decibels. To point out the source of this confusion and, we hope, thereby to avoid it, let us consider the situation represented in Fig. 1.6-1. Here a waveform  $v_i(t) = V_i \cos \omega t$  is applied to the input of a linear amplifier of input impedance  $R_i$ . An output signal  $v_o(t) = V_o \cos(\omega t + \theta)$  then appears across the load resistor  $R_o$ . A real power  $P_i = V_i^2/2R_i$  is supplied to the input, and the real power delivered to the load is  $P_o = V_o^2/2R_o$ . The real power gain  $P_o/P_i$  of the amplifier expressed in decibels is

$$K_{\text{real}} = 10 \log \frac{V_o^2/2R_o}{V_i^2/2R_i} \quad (1.6-3)$$

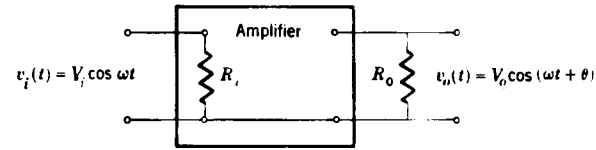


Figure 1.6-1 An amplifier of input impedance  $R_i$  with load  $R_o$ .

If it should happen that  $R_i = R_o$ , then  $K_{\text{real}}$  may be written as in Eq. (1.6-2).

$$K_{\text{real}} = 20 \log \frac{V_o}{V_i} \quad (1.6-4)$$

But if  $R_i \neq R_o$ , then Eq. (1.6-4) does not apply. On the other hand, if we calculate the normalized power gain, then we have

$$K_{\text{norm}} = 10 \log \frac{V_o^2/2}{V_i^2/2} = 20 \log \frac{V_o}{V_i} \quad (1.6-5)$$

So far as the normalized power gain is concerned, the impedances  $R_i$  and  $R_o$  in Fig. 1.6-1 are *absolutely irrelevant*. If it should happen that  $R_i = R_o$ , then  $K_{\text{real}} = K_{\text{norm}}$ , but otherwise they would be different.

## 1.7 NORMALIZED POWER IN A FOURIER EXPANSION

Let us consider two typical terms of the Fourier expansion of Eq. (1.1-5). If we take, say, the fundamental and first harmonic, we have

$$v'(t) = C_1 \cos\left(\frac{2\pi t}{T_0} - \phi_1\right) + C_2 \cos\left(\frac{4\pi t}{T_0} - \phi_2\right) \quad (1.7-1)$$

To calculate the normalized power  $S'$  of  $v'(t)$ , we must square  $v'(t)$  and evaluate

$$S' = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} [v'(t)]^2 dt \quad (1.7-2)$$

When we square  $v'(t)$ , we get the square of the first term, the square of the second term, and then the cross-product term. However, the two cosine functions in Eq. (1.7-1) are *orthogonal*. That is, when their product is integrated over a complete period, the result is zero. Hence in evaluating the normalized power, we find no term corresponding to this cross product. We find actually that  $S'$  is given by

$$S' = \frac{C_1^2}{2} + \frac{C_2^2}{2} \quad (1.7-3)$$

By extension it is apparent that the normalized power associated with the entire Fourier series is

$$S = C_0^2 + \sum_{n=1}^{\infty} \frac{C_n^2}{2} \quad (1.7-4)$$

Hence we observe that because of the orthogonality of the sinusoids used in a Fourier expansion, the total normalized power is the sum of the normalized power due to each term in the series separately. If we write a waveform as a sum of terms which are not orthogonal, this very simple and useful result will not apply. We may note here also that, in terms of the  $A$ 's and  $B$ 's of the Fourier representation of Eq. (1.1-1), the normalized power is

$$S = A_0^2 + \sum_{n=1}^{\infty} \frac{A_n^2}{2} + \sum_{n=1}^{\infty} \frac{B_n^2}{2} \quad (1.7-5)$$

It is to be observed that power and normalized power are to be associated with a *real* waveform and not with a *complex* waveform. Thus suppose we have a term  $A_n \cos(2\pi nt/T_0)$  in a Fourier series. Then the normalized power contributed by this term is  $A_n^2/2$  quite independently of all other terms. And this normalized power comes from averaging, over time, the product of the term  $A_n \cos(2\pi nt/T_0)$  by *itself*. On the other hand, in the complex Fourier representation of Eq. (1.2-1) we have terms of the form  $V_n e^{j2\pi nt/T_0}$ . The average value of the square of such a term is zero. We find as a matter of fact that the contributions to normalized power come from product terms

$$V_n e^{j2\pi nt/T_0} V_{-n} e^{-j2\pi nt/T_0} = V_n V_{-n} = V_n V_n^* \quad (1.7-6)$$

The total normalized power is

$$S = \sum_{n=-\infty}^{+\infty} V_n V_n^* \quad (1.7-7)$$

Thus, in the complex representation, the power associated with a particular *real* frequency  $n/T_0 = nf_0$  ( $f_0$  is the fundamental frequency) is associated neither with the spectral component at  $nf_0$  nor with the component at  $-nf_0$ , but rather with the *combination* of spectral components, one in the positive-frequency range and one in the negative-frequency range. This power is

$$V_n V_n^* + V_{-n} V_{-n}^* = 2V_n V_n^* \quad (1.7-8)$$

It is nonetheless a procedure of great convenience to associate one-half of the power in this combination of spectral components (that is,  $V_n V_n^*$ ) with the frequency  $nf_0$  and the other half with the frequency  $-nf_0$ . Such a procedure will always be valid provided that we are careful to use the procedure to calculate only the *total* power associated with frequencies  $nf_0$  and  $-nf_0$ . Thus we may say that the power associated with the spectral component at  $nf_0$  is  $V_n V_n^*$  and the power associated with the spectral component at  $-nf_0$  is similarly  $V_{-n} V_{-n}^*$  ( $= V_n V_n^*$ ). If we use these associations only to arrive at the result that the total power is  $2V_n V_n^*$ , we shall make no error.

In correspondence with the one-sided and two-sided spectral amplitude pattern of Fig. 1.1-1 we may construct one-sided and two-sided spectral (normalized) power diagrams. A two-sided power spectral diagram is shown in Fig. 1.7-1. The vertical axis is labeled  $S_n$ , the power associated with each spectral

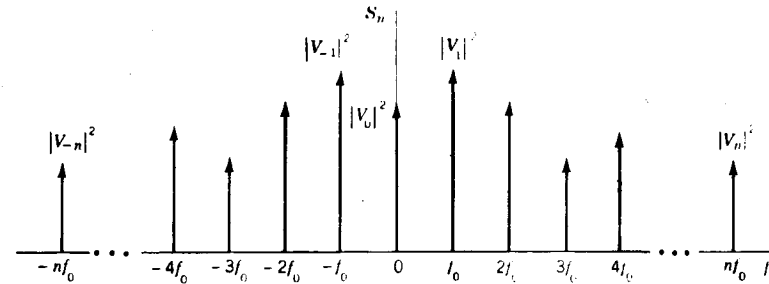


Figure 1.7-1 A two-sided power spectrum.

component. The height of each vertical line is  $|V_n|^2$ . Because of its greater convenience and because it lends a measure of systemization to very many calculations, we shall use the two-sided amplitude and power spectral pattern exclusively throughout this text.

We shall similarly use a two-sided representation to specify the transmission characteristics of filters. Thus, suppose we have a low-pass filter which transmits without attenuation all spectral components up to a frequency  $f_M$  and transmits nothing at a higher frequency. Then the magnitude of the transfer function will be given as in Fig. 1.7-2. The transfer characteristic of a bandpass filter will be given as in Fig. 1.7-3.

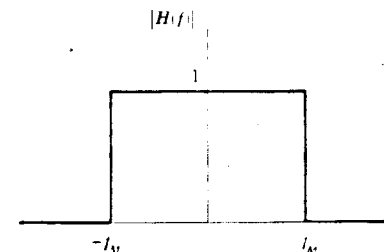


Figure 1.7-2 The transfer characteristic of an idealized low-pass filter.

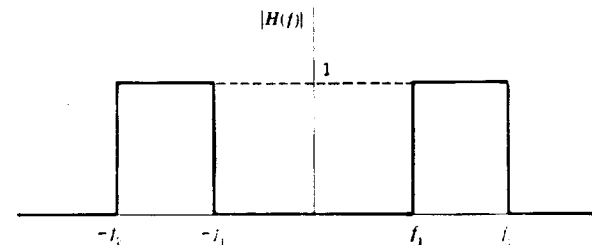


Figure 1.7-3 The transfer characteristic of an idealized bandpass filter with passband from  $f_1$  to  $f_2$ .



## 1.8 POWER SPECTRAL DENSITY

Suppose that, in Fig. 1.7-1 where  $S_n$  is given for each spectral component, we start at  $f = -\infty$  and then, moving in the positive-frequency direction, we add the normalized powers contributed by each power spectral line up to the frequency  $f$ . This sum is  $S(f)$ , a function of frequency.  $S(f)$  typically will have the appearance shown in Fig. 1.8-1. It does not change as  $f$  goes from one spectral line to another, but jumps abruptly as the normalized power of each spectral line is added. Now let us inquire about the normalized power at the frequency  $f$  in a range  $df$ . This quantity of normalized power  $dS(f)$  would be written

$$dS(f) = \frac{dS(f)}{df} df \quad (1.8-1)$$

The quantity  $dS(f)/df$  is called the (normalized) power spectral density  $G(f)$ ; thus

$$G(f) \equiv \frac{dS(f)}{df} \quad (1.8-2)$$

The power in the range  $df$  at  $f$  is  $G(f) df$ . The power in the positive-frequency range  $f_1$  to  $f_2$  is

$$S(f_1 \leq f \leq f_2) = \int_{f_1}^{f_2} G(f) df \quad (1.8-3)$$

The power in the negative-frequency range  $-f_2$  to  $-f_1$  is

$$S(-f_2 \leq f \leq -f_1) = \int_{-f_2}^{-f_1} G(f) df \quad (1.8-4)$$

The quantities in Eqs. (1.8-3) and (1.8-4) do not have physical significance. However, the total power in the real frequency range  $f_1$  to  $f_2$  does have physical significance, and this power  $S(f_1 \leq |f| \leq f_2)$  is given by

$$S(f_1 \leq |f| \leq f_2) = \int_{-f_2}^{-f_1} G(f) df + \int_{f_1}^{f_2} G(f) df \quad (1.8-5)$$

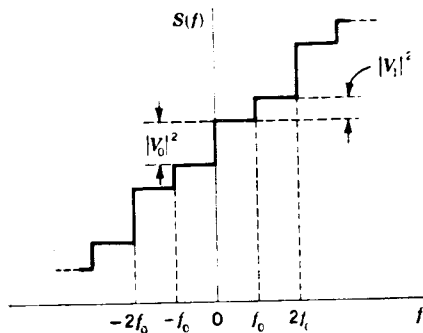


Figure 1.8-1 The sum  $S(f)$  of the normalized power in all spectral components from  $f = -\infty$  to  $f$

To find the power spectral density, we must differentiate  $S(f)$  in Fig. 1.8-1. Between harmonic frequencies we would have  $G(f) = 0$ . At a harmonic frequency,  $G(f)$  would yield an impulse of strength equal to the size of the jump in  $S(f)$ . Thus we would find

$$G(f) = \sum_{n=-\infty}^{\infty} |V_n|^2 \delta(f - nf_0) \quad (1.8-6)$$

If, in plotting  $G(f)$ , we were to represent an impulse by a vertical arrow of height proportional to the impulse strength, then a plot of  $G(f)$  versus  $f$  as given by Eq. (1.8-6) would have exactly the same appearance as the plot of  $S_n$  shown in Fig. 1.7-1.

## 1.9 EFFECT OF TRANSFER FUNCTION ON POWER SPECTRAL DENSITY

Let the input signal  $v_i(t)$  to a filter have a power spectral density  $G_i(f)$ . If  $V_{in}$  are the spectral amplitudes of this input signal, then, using Eq. (1.8-6)

$$G_i(f) = \sum_{n=-\infty}^{\infty} |V_{in}|^2 \delta(f - nf_0) \quad (1.9-1)$$

where, from Eq. (1.2-2),

$$V_{in} = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} v_i(t) e^{-j2\pi nt/T_0} dt \quad (1.9-2)$$

Let the output signal of the filter be  $v_o(t)$ . If  $V_{on}$  are the spectral amplitudes of this output signal, then the corresponding power spectral density is

$$G_o(f) = \sum_{n=-\infty}^{\infty} |V_{on}|^2 \delta(f - nf_0) \quad (1.9-3)$$

where

$$V_{on} = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} v_o(t) e^{-j2\pi nt/T_0} dt \quad (1.9-4)$$

As discussed in Sec. 1.5, if the transfer function of the filter is  $H(f)$ , then the output coefficient  $V_{on}$  is related to the input coefficient by

$$V_{on} = V_{in} H(f = nf_0) \quad (1.9-5)$$

Hence,

$$|V_{on}|^2 = |V_{in}|^2 |H(f = nf_0)|^2 \quad (1.9-6)$$

Substituting Eq. (1.9-6) into Eq. (1.9-3) and comparing the result with Eq. (1.9-1) yields the important result

$$G_o(f) = G_i(f) |H(f)|^2 \quad (1.9-7)$$

Equation (1.9-7) is of the greatest importance since it relates the power spectral density, and hence the power, at one point in a system to the power spectral density at another point in the system. Equation (1.9-7) was derived for the special case of periodic signals; however, it applies to nonperiodic signals and signals represented by random processes (Sec. 2.23) as well.

As a special application of interest of the result given in Eq. (1.9-7), assume that an input signal  $v_i(t)$  with power spectral density  $G_i(f)$  is passed through a differentiator. The differentiator output  $v_o(t)$  is related to the input by

$$v_o(t) = \tau \frac{d}{dt} v_i(t) \quad (1.9-8)$$

where  $\tau$  is a constant. The operation indicated in Eq. (1.9-8) multiplies each spectral component of  $v_i(t)$  by  $j2\pi f\tau = j\omega\tau$ . Hence  $H(f) = j\omega\tau$ , and  $|H(f)|^2 = \omega^2\tau^2$ . Thus from Eq. (1.9-7) the spectral density of the output is

$$G_o(f) = \omega^2\tau^2 G_i(f) \quad (1.9-9)$$

## 1.10 THE FOURIER TRANSFORM

A periodic waveform may be expressed, as we have seen, as a sum of spectral components. These components have finite amplitudes and are separated by finite frequency intervals  $f_0 = 1/T_0$ . The normalized power of the waveform is finite, as is also the normalized energy of the signal in an interval  $T_0$ . Now suppose we increase without limit the period  $T_0$  of the waveform. Thus, say, in Fig. 1.3-1b the pulse centered around  $t = 0$  remains in place, but all other pulses move outward away from  $t = 0$  as  $T_0 \rightarrow \infty$ . Then eventually we would be left with a single-pulse nonperiodic waveform.

As  $T_0 \rightarrow \infty$ , the spacing between spectral components becomes infinitesimal. The frequency of the spectral components, which in the Fourier series was a discontinuous variable with a one-to-one correspondence with the integers, becomes instead a continuous variable. The normalized energy of the nonperiodic waveform remains finite, but, since the waveform is not repeated, its normalized power becomes infinitesimal. The spectral amplitudes similarly become infinitesimal. The Fourier series for the periodic waveform

$$v(t) = \sum_{n=-\infty}^{\infty} V_n e^{j2\pi n f_0 t} \quad (1.10-1)$$

becomes (see Prob. 1.10-1)

$$v(t) = \int_{-\infty}^{\infty} V(f) e^{j2\pi f t} df \quad (1.10-2)$$

The finite spectral amplitudes  $V_n$  are analogous to the infinitesimal spectral

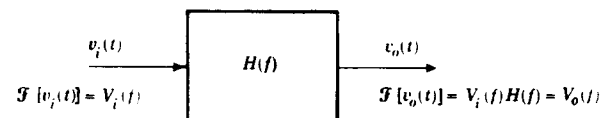


Figure 1.10-1 A waveform  $v_i(t)$  of transform  $V_i(f)$  is transmitted through a network of transfer function  $H(f)$ . The output waveform  $v_o(t)$  has a transform  $V_o(f) = V_i(f)H(f)$ .

amplitudes  $V(f) df$ . The quantity  $V(f)$  is called the *amplitude spectral density* or more generally the *Fourier transform* of  $v(t)$ . The Fourier transform is given by

$$V(f) = \int_{-\infty}^{\infty} v(t) e^{-j2\pi f t} dt \quad (1.10-3)$$

in correspondence with  $V_n$ , which is given by

$$V_n = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} v(t) e^{-j2\pi n f_0 t} dt \quad (1.10-4)$$

Again, in correspondence with Eq. (1.5-8), let  $H(f)$  be the transfer function of a network. If the input signal is  $v_i(t)$ , then the output signal will be  $v_o(t)$  given by

$$v_o(t) = \int_{-\infty}^{\infty} H(f) V(f) e^{j2\pi f t} df \quad (1.10-5)$$

Comparing Eq. (1.10-5) with Eq. (1.10-2), we see that the Fourier transform  $V_o(f) \equiv \mathcal{F}[v_o(t)]$  is related to the transform  $V_i(f)$  of  $v_i(t)$  by

$$\mathcal{F}[v_o(t)] = H(f) \mathcal{F}[v_i(t)] \quad (1.10-6)$$

or

$$V_o(f) = H(f) V_i(f) \quad (1.10-7)$$

as indicated in Fig. 1.10-1.

## 1.11 EXAMPLES OF FOURIER TRANSFORMS

In this section we shall evaluate the transforms of a number of functions both for the sake of review and also because we shall have occasion to refer to the results.

**Example 1.11-1** If  $v(t) = \cos \omega_0 t$ , find  $V(f)$ .

**SOLUTION** The function  $v(t) = \cos \omega_0 t$  is periodic, and therefore has a Fourier series representation as well as a Fourier transform.

The exponential Fourier series representation of  $v(t)$  is

$$v(t) = \frac{1}{2} e^{+j\omega_0 t} + \frac{1}{2} e^{-j\omega_0 t} \quad \omega_0 = \frac{2\pi}{T_0} \quad (1.11-1)$$

Thus

$$V_1 = V_{-1} = \frac{1}{2} \quad (1.11-2)$$

and

$$V_n = 0 \quad n \neq \pm 1 \quad (1.11-3)$$

The Fourier transform  $V(f)$  is found using Eq. (1.10-3):

$$V(f) = \int_{-\infty}^{\infty} \cos \omega_0 t e^{-j2\pi f t} dt = \frac{1}{2} \int_{-\infty}^{\infty} e^{-j2\pi(f-f_0)t} dt + \frac{1}{2} \int_{-\infty}^{\infty} e^{-j2\pi(f+f_0)t} dt \quad (1.11-4a)$$

$$= \frac{1}{2} \delta(f-f_0) + \frac{1}{2} \delta(f+f_0) \quad (1.11-4b)$$

[See Eq. (1.11-22) below.]

From Eqs. (1.11-1) and (1.11-3) we draw the following conclusion: The Fourier transform of a sinusoidal signal (or other periodic signal) consists of impulses located at each harmonic frequency of the signal, i.e., at  $f_n = n/T_0 = n f_0$ . The strength of each impulse is equal to the amplitude of the Fourier coefficient of the exponential series.

**Example 1.11-2** A signal  $m(t)$  is multiplied by a sinusoidal waveform of frequency  $f_c$ . The product signal is

$$v(t) = m(t) \cos 2\pi f_c t \quad (1.11-5)$$

If the Fourier transform of  $m(t)$  is  $M(f)$ , that is,

$$M(f) = \int_{-\infty}^{\infty} m(t) e^{-j2\pi f t} dt \quad (1.11-6)$$

find the Fourier transform of  $v(t)$ .

**SOLUTION** Since

$$m(t) \cos 2\pi f_c t = \frac{1}{2} m(t) e^{j2\pi f_c t} + \frac{1}{2} m(t) e^{-j2\pi f_c t} \quad (1.11-7)$$

then the Fourier transform  $V(f)$  is given by

$$V(f) = \frac{1}{2} \int_{-\infty}^{\infty} m(t) e^{-j2\pi(f+f_c)t} dt + \frac{1}{2} \int_{-\infty}^{\infty} m(t) e^{-j2\pi(f-f_c)t} dt \quad (1.11-8)$$

Comparing Eq. (1.11-8) with Eq. (1.11-6), we have the result that

$$V(f) = \frac{1}{2} M(f+f_c) + \frac{1}{2} M(f-f_c) \quad (1.11-9)$$

The relationship of the transform  $M(f)$  of  $m(t)$  to the transform  $V(f)$  of  $m(t) \cos 2\pi f_c t$  is illustrated in Fig. 1.11-1a. In Fig. 1.11-1b we see the spectral pattern of  $M(f)$  replaced by two patterns of the same form. One is shifted to the right and one to the left, each by amount  $f_c$ . Further, the amplitudes of each of these two spectral patterns is one-half the amplitude of the spectral pattern  $M(f)$ .

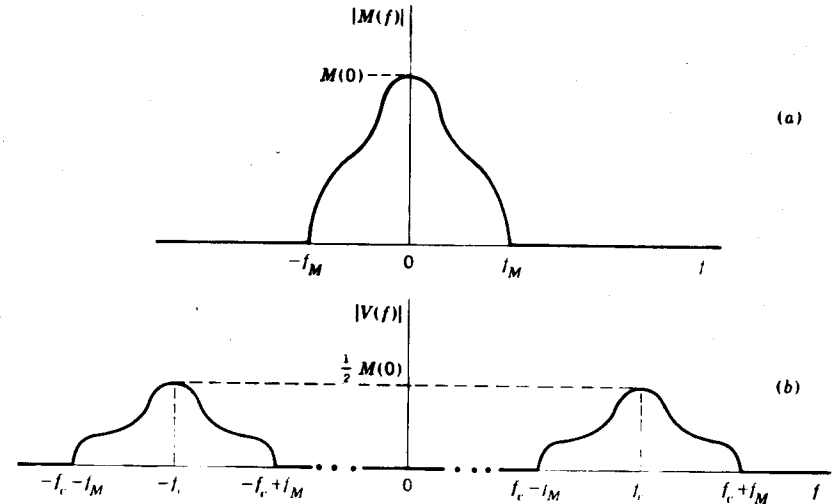


Figure 1.11-1 (a) The amplitude spectrum of a waveform with no spectral component beyond  $f_M$ . (b) The amplitude spectrum of the waveform in (a) multiplied by  $\cos 2\pi f_c t$ .

A case of special interest arises when the waveform  $m(t)$  is itself sinusoidal. Thus assume

$$m(t) = m \cos 2\pi f_m t \quad (1.11-10)$$

where  $m$  is a constant. We then find that  $V(f)$  is given by

$$V(f) = \frac{m}{4} \delta(f+f_c+f_m) + \frac{m}{4} \delta(f+f_c-f_m) + \frac{m}{4} \delta(f-f_c+f_m) + \frac{m}{4} \delta(f-f_c-f_m) \quad (1.11-11)$$

This spectral pattern is shown in Fig. 1.11-2. Observe that the pattern has four spectral lines corresponding to two real frequencies  $f_c + f_m$  and  $f_c - f_m$ .

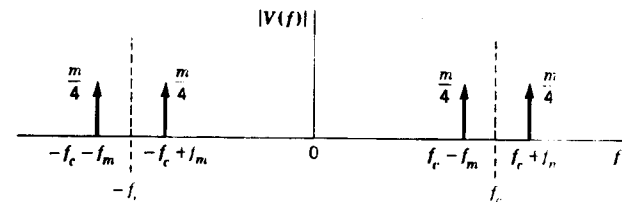


Figure 1.11-2 The two-sided amplitude spectrum of the product waveform  $v(t) = m \cos 2\pi f_c t$ .

The waveform itself is given by

$$v(t) = \frac{m}{4} [e^{j2\pi(f_c + f_m)t} + e^{-j2\pi(f_c + f_m)t}] + \frac{m}{4} [e^{j2\pi(f_c - f_m)t} + e^{-j2\pi(f_c - f_m)t}] \quad (1.11-12a)$$

$$= \frac{m}{4} [\cos 2\pi(f_c + f_m)t + \cos 2\pi(f_c - f_m)t] \quad (1.11-12b)$$

**Example 1.11-3** A pulse of amplitude  $A$  extends from  $t = -\tau/2$  to  $t = +\tau/2$ . Find its Fourier transform  $V(f)$ . Consider also the Fourier series for a periodic sequence of such pulses separated by intervals  $T_0$ . Compare the Fourier series coefficients  $V_n$  with the transform in the limit as  $T_0 \rightarrow \infty$ .

**SOLUTION** We have directly that

$$V(f) = \int_{-\tau/2}^{\tau/2} A e^{-j2\pi f t} dt = A\tau \frac{\sin \pi f \tau}{\pi f \tau} \quad (1.11-13)$$

The Fourier series coefficients of the periodic pulse train are given by Eq. (1.3-12b) as

$$V_n = \frac{A\tau}{T_0} \frac{\sin(n\pi\tau/T_0)}{n\pi\tau/T_0} \quad (1.11-14)$$

The fundamental frequency in the Fourier series is  $f_0 = 1/T_0$ . We shall set  $f_0 \equiv \Delta f$  in order to emphasize that  $f_0 \equiv \Delta f$  is the frequency interval between spectral lines in the Fourier series. Hence, since  $1/T_0 = \Delta f$ , we may rewrite Eq. (1.11-14) as

$$V_n = A\tau \frac{\sin(\pi n \Delta f \tau)}{\pi n \Delta f \tau} \Delta f \quad (1.11-15)$$

In the limit, as  $T_0 \rightarrow \infty$ ,  $\Delta f \rightarrow 0$ . We then may replace  $\Delta f$  by  $df$  and replace  $n \Delta f$  by a continuous variable  $f$ . Equation (1.11-15) then becomes

$$\lim_{\Delta f \rightarrow 0} V_n = A\tau \frac{\sin \pi f \tau}{\pi f \tau} df \quad (1.11-16)$$

Comparing this result with Eq. (1.11-13), we do indeed note that

$$V(f) = \lim_{\Delta f \rightarrow 0} \frac{V_n}{\Delta f} \quad (1.11-17)$$

Thus we confirm our earlier interpretation of  $V(f)$  as an *amplitude spectral density*.

#### Example 1.11-4

- (a) Find the Fourier transform of  $\delta(t)$ , an impulse of unit strength.  
 (b) Given a network whose transfer function is  $H(f)$ . An impulse  $\delta(t)$  is applied at the input. Show that the response  $v_o(t) \equiv h(t)$  at the output is the inverse transform of  $H(f)$ , that is, show that  $h(t) = \mathcal{F}^{-1}[H(f)]$ .

**SOLUTION**

- (a) The impulse  $\delta(t) = 0$  except at  $t = 0$  and, further, has the property that

$$\int_{-\infty}^{\infty} \delta(t) dt = 1 \quad (1.11-18)$$

Hence 
$$V(f) = \int_{-\infty}^{\infty} \delta(t) e^{-j2\pi f t} dt = 1 \quad (1.11-19)$$

Thus the spectral components of  $\delta(t)$  extend with uniform amplitude and phase over the entire frequency domain.

- (b) Using the result given in Eq. (1.10-7), we find that the transform of the output  $v_o(t) \equiv h(t)$  is  $V_o(f)$  given by

$$V_o(f) = 1 \times H(f) \quad (1.11-20)$$

since the transform of  $\delta(t)$ ,  $\mathcal{F}[\delta(t)] = 1$ . Hence the inverse transform of  $V_o(f)$ , which is the function  $h(t)$ , is also the inverse transform of  $H(f)$ . Specifically, for an impulse input, the output is

$$h(t) = \int_{-\infty}^{\infty} H(f) e^{j2\pi f t} df \quad (1.11-21)$$

We may use the result given in Eq. (1.11-21) to arrive at a useful representation of  $\delta(t)$  itself. If  $H(f) = 1$ , then the response  $h(t)$  to an impulse  $\delta(t)$  is the impulse itself. Hence, setting  $H(f) = 1$  in Eq. (1.11-21), we find

$$\delta(t) = \int_{-\infty}^{\infty} e^{j2\pi f t} df = \int_{-\infty}^{\infty} e^{-j2\pi f t} df \quad (1.11-22)$$

## 1.12 CONVOLUTION

Suppose that  $v_1(t)$  has the Fourier transform  $V_1(f)$  and  $v_2(t)$  has the transform  $V_2(f)$ . What then is the waveform  $v(t)$  whose transform is the product  $V_1(f)V_2(f)$ ? This question arises frequently in spectral analysis and is answered by the *convolution theorem*, which says that

$$v(t) = \int_{-\infty}^{\infty} v_1(\tau) v_2(t - \tau) d\tau \quad (1.12-1)$$

or equivalently

$$v(t) = \int_{-\infty}^{\infty} v_2(\tau) v_1(t - \tau) d\tau \quad (1.12-2)$$

The integrals in Eq. (1.12-1) or (1.12-2) are called *convolution integrals*, and the process of evaluating  $v(t)$  through these integrals is referred to as *taking the convolution* of the functions  $v_1(t)$  and  $v_2(t)$ .

To prove the theorem, we begin by writing

$$v(t) = \mathcal{F}^{-1}[V_1(f)V_2(f)] \quad (1.12-3a)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} V_1(f)V_2(f)e^{j\omega t} d\omega \quad (1.12-3b)$$

By definition we have

$$V_1(f) = \int_{-\infty}^{\infty} v_1(\tau)e^{-j\omega\tau} d\tau \quad (1.12-4)$$

Substituting  $V_1(f)$  as given by Eq. (1.12-4) into the integrand of Eq. (1.12-3b), we have

$$v(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} v_1(\tau)e^{-j\omega\tau} d\tau V_2(f)e^{j\omega t} d\omega \quad (1.12-5)$$

Interchanging the order of integration, we find

$$v(t) = \int_{-\infty}^{\infty} v_1(\tau) \left[ \frac{1}{2\pi} \int_{-\infty}^{\infty} V_2(f)e^{j\omega(t-\tau)} d\omega \right] d\tau \quad (1.12-6)$$

We recognize that the expression in brackets in Eq. (1.12-6) is  $v_2(t - \tau)$ , so that finally

$$v(t) = \int_{-\infty}^{\infty} v_1(\tau)v_2(t - \tau) d\tau \quad (1.12-7)$$

Examples of the use of the convolution integral are given in Probs. 1.12-2 and 1.12-3. These problems will also serve to recall to the reader the relevance of the term *convolution*.

A special case of the convolution theorem and one of very great utility is arrived at through the following considerations. Suppose a waveform  $v_i(t)$  whose transform is  $V_i(f)$  is applied to a linear network with transfer function  $H(f)$ . The transform of the output waveform is  $V_o(f)H(f)$ . What then is the waveform of  $v_o(t)$ ?

In Eq. (1.12-7) we identify  $v_1(\tau)$  with  $v_i(\tau)$  and  $v_2(t)$  with the inverse transform of  $H(f)$ . But we have seen [in Eq. (1.11-21)] that the inverse transform of  $H(f)$  is  $h(t)$ , the impulse response of the network. Hence Eq. (1.12-7) becomes

$$v_o(t) = \int_{-\infty}^{\infty} v_i(\tau)h(t - \tau) d\tau \quad (1.12-8)$$

in which the output  $v_o(t)$  is expressed in terms of the input  $v_i(t)$  and the impulse response of the network.

### 1.13 PARSEVAL'S THEOREM

We saw that for periodic waveforms we may express the normalized power as a summation of powers due to individual spectral components. We also found for periodic signals that it was appropriate to introduce the concept of *power spectral density*. Let us keep in mind that we may make the transition from the periodic to the nonperiodic waveform by allowing the period of the periodic waveform to approach infinity. A nonperiodic waveform, so generated, has a finite *normalized energy*, while the normalized power approaches zero. We may therefore expect that for a nonperiodic waveform the energy may be written as a continuous summation (integral) of energies due to individual spectral components in a continuous distribution. Similarly we should expect that with such nonperiodic waveforms it should be possible to introduce an *energy spectral density*.

The normalized energy of a periodic waveform  $v(t)$  in a period  $T_0$  is

$$E = \int_{-T_0/2}^{T_0/2} [v(t)]^2 dt \quad (1.13-1)$$

From Eq. (1.7-7) we may write  $E$  as

$$E = T_0 S = T_0 \sum_{n=-\infty}^{+\infty} V_n V_n^* \quad (1.13-2)$$

Again, as in the illustrative Example 1.11-3, we let  $\Delta f \equiv 1/T_0 = f_0$ , where  $f_0$  is the fundamental frequency, that is,  $\Delta f$  is the spacing between harmonics. Then we have

$$V_n V_n^* = \frac{V_n}{\Delta f} \frac{V_n^*}{\Delta f} (\Delta f)^2 \quad (1.13-3)$$

and Eq. (1.13-2) may be written

$$E = \sum_{n=-\infty}^{\infty} \frac{V_n}{\Delta f} \frac{V_n^*}{\Delta f} \Delta f \quad (1.13-4)$$

In the limit, as  $\Delta f \rightarrow 0$ , we replace  $\Delta f$  by  $df$ , we replace  $V_n/\Delta f$  by the transform  $V(f)$  [see Eq. (1.11-17)], and the summation by an integral. Equation (1.13-4) then becomes

$$E = \int_{-\infty}^{\infty} V(f)V^*(f) df = \int_{-\infty}^{\infty} |V(f)|^2 df = \int_{-\infty}^{\infty} [v(t)]^2 dt \quad (1.13-5)$$

This equation expresses *Parseval's theorem*. Parseval's theorem is the extension to the nonperiodic case of Eq. (1.7-7) which applies for periodic waveforms. Both results simply express the fact that the power (periodic case) or the energy (nonperiodic case) may be written as the superposition of power or energy due to individual spectral components separately. The validity of these results depends on the fact that the spectral components are orthogonal. In the periodic case the spectral components are orthogonal over the interval  $T_0$ . In the nonperiodic case

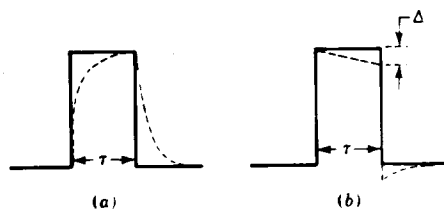


Figure 1.15-2 (a) A rectangular pulse (solid) and the response (dashed) of a low-pass RC circuit. The rise time  $t_r \approx 0.35\tau$ . (b) The response of a high-pass RC circuit for  $f_1\tau = 0.02$ .

The essential distortion introduced by the frequency discrimination of this network is that the output does not sustain a constant voltage level when the input is constant. Instead, the output begins immediately to decay toward zero, which is its asymptotic limit. From Eq. (1.15-7) we have

$$\frac{1}{v_o} \frac{dv_o}{dt} = -\frac{1}{RC} = -2\pi f_1 \quad (1.15-8)$$

Thus the low-frequency cutoff  $f_1$  alone determines the percentage drop in voltage per unit time. Again the importance of Eq. (1.15-8) is that, at least as a reasonable approximation, it applies to high-pass networks quite generally, even when the network is very much more complicated than the simple RC circuit.

If the input to the RC high-pass circuit is a pulse of duration  $\tau$ , then the output has the waveshape shown in Fig. 1.15-2b. The output exhibits a tilt and an undershoot. As a rule of thumb, we may assume that the pulse is reasonably faithfully reproduced if the tilt  $\Delta$  is no more than  $0.1V_o$ . Correspondingly, this condition requires that  $f_1$  be no higher than given by the condition

$$f_1\tau \approx 0.02 \quad (1.15-9)$$

## 1.16 CORRELATION BETWEEN WAVEFORMS

The correlation between waveforms is a measure of the similarity or relatedness between the waveforms. Suppose that we have waveforms  $v_1(t)$  and  $v_2(t)$ , not necessarily periodic nor confined to a finite time interval. Then the correlation between them, or more precisely the *average cross correlation* between  $v_1(t)$  and  $v_2(t)$ , is  $R_{12}(\tau)$  defined as

$$R_{12}(\tau) \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} v_1(t)v_2(t+\tau) dt \quad (1.16-1)$$

If  $v_1(t)$  and  $v_2(t)$  are periodic with the same fundamental period  $T_0$ , then the average cross correlation is

$$R_{12}(\tau) = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} v_1(t)v_2(t+\tau) dt \quad (1.16-2)$$

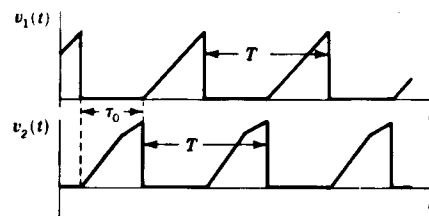


Figure 1.16-1 Two related waveforms. The timing is such that the product  $v_1(t)v_2(t) = 0$ .

If  $v_1(t)$  and  $v_2(t)$  are waveforms of finite energy (for example, nonperiodic pulse-type waveforms), then the cross correlation is defined as

$$R_{12}(\tau) = \int_{-\infty}^{\infty} v_1(t)v_2(t+\tau) dt \quad (1.16-3)$$

The need for introducing the parameter  $\tau$  in the definition of cross correlation may be seen by the example illustrated in Fig. 1.16-1. Here the two waveforms, while different, are obviously related. They have the same period and nearly the same form. However, the integral of the product  $v_1(t)v_2(t)$  is zero since at all times one or the other function is zero. The function  $v_2(t+\tau)$  is the function  $v_2(t)$  shifted to the left by amount  $\tau$ . It is clear from the figure that, while  $R_{12}(0) = 0$ ,  $R_{12}(\tau)$  will increase as  $\tau$  increases from zero, becoming a maximum when  $\tau = \tau_0$ . Thus  $\tau$  is a "searching" or "scanning" parameter which may be adjusted to a proper time shift to reveal, to the maximum extent possible, the relatedness or correlation between the functions. The term *coherence* is sometimes used as a synonym for correlation. Functions for which  $R_{12}(\tau) = 0$  for all  $\tau$  are described as being *uncorrelated* or *noncoherent*.

In scanning to see the extent of the correlation between functions, it is necessary to specify which function is being shifted. In general,  $R_{12}(\tau)$  is not equal to  $R_{21}(\tau)$ . It is readily verified (Prob. 1.16-2) that

$$R_{21}(\tau) \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} v_1(t+\tau)v_2(t) dt = R_{12}(-\tau) \quad (1.16-4)$$

with identical results for periodic waveforms or waveforms of finite energy.

## 1.17 POWER AND CROSS CORRELATION

Let  $v_1(t)$  and  $v_2(t)$  be waveforms which are not periodic nor confined to a finite time interval. Suppose that the normalized power of  $v_1(t)$  is  $S_1$  and the normalized power of  $v_2(t)$  is  $S_2$ . What, then, is the normalized power of  $v_1(t) + v_2(t)$ ? Or,

more generally, what is the normalized power  $S_{12}$  of  $v_1(t) + v_2(t + \tau)$ ? We have

$$S_{12} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} [v_1(t) + v_2(t + \tau)]^2 dt \quad (1.17-1a)$$

$$= \lim_{T \rightarrow \infty} \frac{1}{T} \left\{ \int_{-T/2}^{T/2} v_1^2(t) dt + \int_{-T/2}^{T/2} [v_2(t + \tau)]^2 dt + 2 \int_{-T/2}^{T/2} v_1(t)v_2(t + \tau) dt \right\} \quad (1.17-1b)$$

$$= S_1 + S_2 + 2R_{12}(\tau) \quad (1.17-1c)$$

In writing Eq. (1.17-1c), we have taken account of the fact that the normalized power of  $v_2(t + \tau)$  is the same as the normalized power of  $v_2(t)$ . For, since the integration in Eq. (1.17-1b) extends eventually over the entire time axis, a time shift in  $v_2$  will clearly not affect the value of the integral.

From Eq. (1.17-1c) we have the important result that if two waveforms are uncorrelated, that is,  $R_{12}(\tau) = 0$  for all  $\tau$ , then no matter how these waveforms are time-shifted with respect to one another, the normalized power due to the superposition of the waveforms is the sum of the powers due to the waveforms individually. Similarly if a waveform is the sum of any number of mutually uncorrelated waveforms, the normalized power is the sum of the individual powers. It is readily verified that the same result applies for periodic waveforms. For finite energy waveforms, the result applies to the normalized energy.

Suppose that two waveforms  $v_1(t)$  and  $v_2(t)$  are uncorrelated. If dc components  $V_1$  and  $V_2$  are added to the waveforms, then the waveforms  $v_1(t) = v_1'(t) + V_1$  and  $v_2(t) = v_2'(t) + V_2$  will be correlated with correlation  $R_{12}(\tau) = V_1 V_2$ . In most applications where the correlations between waveforms is of concern, there is rarely any interest in the dc component. It is customary, then, to continue to refer to waveforms as being uncorrelated if the only source of the correlation is the dc components.

## 1.18 AUTOCORRELATION

The correlation of a function with itself is called the *autocorrelation*. Thus with  $v_1(t) = v_2(t)$ ,  $R_{12}(\tau)$  becomes  $R(\tau)$  given, in the general case, by

$$R(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} v(t)v(t + \tau) dt \quad (1.18-1)$$

A number of the properties of  $R(\tau)$  are listed in the following:

$$(a) \quad R(0) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} [v(t)]^2 dt = S \quad (1.18-2)$$

That is, the autocorrelation for  $\tau = 0$  is the average power  $S$  of the waveform.

$$(b) \quad R(0) \geq R(\tau) \quad (1.18-3)$$

This result is rather intuitively obvious since we would surely expect that similarity between  $v(t)$  and  $v(t + \tau)$  be a maximum when  $\tau = 0$ . The student is guided through a more formal proof in Prob. 1.18-1.

$$(c) \quad R(\tau) = R(-\tau) \quad (1.18-4)$$

Thus the autocorrelation function is an even function of  $\tau$ . To prove Eq. (1.18-4), assume that the axis  $t = 0$  is moved in the negative  $t$  direction by amount  $\tau$ . Then the integrand in Eq. (1.18-1) would become  $v(t - \tau)v(t)$ , and  $R(\tau)$  would become  $R(-\tau)$ . Since, however, the integration eventually extends from  $-\infty$  to  $\infty$ , such a shift in time axis can have no effect on the value of the integral. Thus  $R(\tau) = R(-\tau)$ .

The three characteristics given in Eqs. (1.18-2) to (1.18-4) are features not only of  $R(\tau)$  defined by Eq. (1.18-1) but also for  $R(\tau)$  as defined by Eqs. (1.16-2) and (1.16-3) for the periodic case and the non-periodic case of finite energy. In the latter case, of course,  $R(0) = E$ , the energy rather than the power.

## 1.19 AUTOCORRELATION OF A PERIODIC WAVEFORM

When the function is periodic, we may write

$$v(t) = \sum_{n=-\infty}^{\infty} V_n e^{j2\pi n t / T_0} \quad (1.19-1)$$

and, using the correlation integral in the form of Eq. (1.16-2), we have

$$R(\tau) = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} \left( \sum_{m=-\infty}^{\infty} V_m e^{j2\pi m t / T_0} \right) \left( \sum_{n=-\infty}^{\infty} V_n e^{j2\pi n (t + \tau) / T_0} \right) dt \quad (1.19-2)$$

The order of integration and summation may be interchanged in Eq. (1.19-2). If we do so, we shall be left with a double summation over  $m$  and  $n$  of terms  $I_{m,n}$  given by

$$I_{m,n} = \frac{1}{T_0} e^{j2\pi n \tau / T_0} \int_{-T_0/2}^{T_0/2} V_m V_n e^{j2\pi (m+n)t / T_0} dt \quad (1.19-3a)$$

$$= V_m V_n e^{j2\pi n \tau / T_0} \frac{[\sin \pi(m+n)]}{\pi(m+n)} \quad (1.19-3b)$$

Since  $m$  and  $n$  are integers, we see from Eq. (1.19-3b) that  $I_{m,n} = 0$  except when  $m = -n$  or  $m + n = 0$ . To evaluate  $I_{m,n}$  in this latter case, we return to Eq. (1.19-3a) and find

$$I_{m,n} = I_{-n,n} = V_n V_{-n} e^{j2\pi n \tau / T_0} \quad (1.19-4)$$

Finally, then,

$$R(\tau) = \sum_{n=-\infty}^{\infty} V_n V_{-n} e^{j2\pi n\tau/T_0} = \sum_{n=-\infty}^{\infty} |V_n|^2 e^{j2\pi n\tau/T_0} \quad (1.19-5a)$$

$$= |V_0|^2 + 2 \sum_{n=1}^{\infty} |V_n|^2 \cos 2\pi n \frac{\tau}{T_0} \quad (1.19-5b)$$

We note from Eq. (1.19-5b) that  $R(\tau) = R(-\tau)$  as anticipated, and we note as well that for this case of a periodic waveform the correlation  $R(\tau)$  is also periodic with the same fundamental period  $T_0$ .

We shall now relate the correlation function  $R(\tau)$  of a periodic waveform to its power spectral density. For this purpose we compute the Fourier transform of  $R(\tau)$ . We find, using  $R(\tau)$  as in Eq. (1.19-5a), that

$$\mathcal{F}[R(\tau)] = \int_{-\infty}^{\infty} \left( \sum_{n=-\infty}^{\infty} |V_n|^2 e^{j2\pi n\tau/T_0} \right) e^{-j2\pi f\tau} d\tau \quad (1.19-6)$$

Interchanging the order of integration and summation yields

$$\mathcal{F}[R(\tau)] = \sum_{n=-\infty}^{\infty} |V_n|^2 \int_{-\infty}^{\infty} e^{-j2\pi(f - n/T_0)\tau} d\tau \quad (1.19-7)$$

Using Eq. (1.11-22), we may write Eq. (1.19-7) as

$$\mathcal{F}[R(\tau)] = \sum_{n=-\infty}^{\infty} |V_n|^2 \delta\left(f - \frac{n}{T_0}\right) \quad (1.19-8)$$

Comparing Eq. (1.19-8) with Eq. (1.8-6), we have the interesting result that for a periodic waveform

$$G(f) = \mathcal{F}[R(\tau)] \quad (1.19-9)$$

and, of course, conversely

$$R(\tau) = \mathcal{F}^{-1}[G(f)] \quad (1.19-10)$$

Expressed in words, we have the following: *The power spectral density and the correlation function of a periodic waveform are a Fourier transform pair.*

## 1.20 AUTOCORRELATION OF NONPERIODIC WAVEFORM OF FINITE ENERGY

For pulse-type waveforms of finite energy there is a relationship between the correlation function of Eq. (1.18-1) and the energy spectral density which corresponds to the relationship given in Eq. (1.19-9) for the periodic waveform. This relationship is that the correlation function  $R(\tau)$  and the *energy* spectral density are a Fourier transform pair. This result is established as follows.

We use the convolution theorem. We combine Eqs. (1.12-1) and (1.12-3) for the case where the waveforms  $v_1(t)$  and  $v_2(t)$  are the same waveforms, that is,  $v_1(t) = v_2(t) = v(t)$ , and get

$$\mathcal{F}^{-1}[V(f)V^*(f)] = \int_{-\infty}^{\infty} v(\tau)v(t-\tau) d\tau \quad (1.20-1)$$

Since  $V(-f) = V^*(f) = \mathcal{F}[v(-t)]$ , Eq. (1.20-1) may be written

$$\mathcal{F}^{-1}[V(f)V^*(f)] = \mathcal{F}^{-1}[|V(f)|^2] = \int_{-\infty}^{\infty} v(\tau)v(\tau-t) d\tau \quad (1.20-2)$$

The integral in Eq. (1.20-2) is a function of  $t$ , and hence this equation expresses  $\mathcal{F}^{-1}[V(f)V^*(f)]$  as a function of  $t$ . If we want to express  $\mathcal{F}^{-1}[V(f)V^*(f)]$  as a function of  $\tau$  without changing the form of the function, we need but to interchange  $t$  and  $\tau$ . We then have

$$\mathcal{F}^{-1}[V(f)V^*(f)] = \int_{-\infty}^{\infty} v(t)v(t-\tau) d\tau \quad (1.20-3)$$

The integral in Eq. (1.20-3) is precisely  $R(\tau)$ , and thus

$$\mathcal{F}[R(\tau)] = V(f)V^*(f) = |V(f)|^2 \quad (1.20-4)$$

which verifies that  $R(\tau)$  and the energy spectral density  $|V(f)|^2$  are Fourier transform pairs.

## 1.21 AUTOCORRELATION OF OTHER WAVEFORMS

In the preceding sections we discussed the relationship between the autocorrelation function and power or energy spectral density of *deterministic* waveforms. We use the term "deterministic" to indicate that at least, in principle, it is possible to write a function which specifies the value of the function at all times. For such deterministic waveforms, the availability of an autocorrelation function is of no particularly great value. The autocorrelation function does not include within itself complete information about the function. Thus we note that the autocorrelation function is related only to the amplitudes and not to the phases of the spectral components of the waveform. The waveform cannot be reconstructed from a knowledge of the autocorrelation functions. Any characteristic of a deterministic waveform which may be calculated with the aid of the autocorrelation function may be calculated by direct means at least as conveniently.

On the other hand, in the study of communication systems we encounter waveforms which are not deterministic but are instead random and unpredictable in nature. Such waveforms are discussed in Chap. 2. There we shall find that for such random waveforms no explicit function of time can be written. The waveforms must be described in statistical and probabilistic terms. It is in connection with such waveforms that the concepts of correlation and autocorrelation find their true usefulness. Specifically, it turns out that even for such random wave-



forms, the autocorrelation function and power spectral density are a Fourier transform pair. The proof<sup>3</sup> that such is the case is formidable and will not be undertaken here.

## 1.22 EXPANSIONS IN ORTHOGONAL FUNCTIONS

Let us consider a set of functions  $g_1(x), g_2(x), \dots, g_n(x) \dots$ , defined over the interval  $x_1 \leq x < x_2$  and which are related to one another in the very special way that any two different ones of the set satisfy the condition

$$\int_{x_1}^{x_2} g_i(x) g_j(x) dx = 0 \quad (1.22-1)$$

That is, when we multiply two different functions and then integrate over the interval from  $x_1$  to  $x_2$  the result is zero. A set of functions which has this property is described as being *orthogonal over the interval from  $x_1$  to  $x_2$* . The term "orthogonal" is employed here in correspondence to a similar situation which is encountered in dealing with vectors. The *scalar* product of two vectors  $\mathbf{V}_i$  and  $\mathbf{V}_j$  (also referred to as the *dot* product or as the *inner* product) is a scalar quantity  $V_{ij}$  defined as

$$V_{ij} = |\mathbf{V}_i| |\mathbf{V}_j| \cos(\mathbf{V}_i, \mathbf{V}_j) = V_{ji} \quad (1.22-2)$$

In Eq. (1.22-2)  $|\mathbf{V}_i|$  and  $|\mathbf{V}_j|$  are the magnitudes of the respective vectors and  $\cos(\mathbf{V}_i, \mathbf{V}_j)$  is the cosine of the angle between the vectors. If it should turn out that  $V_{ij} = 0$  then (ignoring the trivial cases in which  $\mathbf{V}_i = 0$  or  $\mathbf{V}_j = 0$ )  $\cos(\mathbf{V}_i, \mathbf{V}_j)$  must be zero and correspondingly it means that the vectors  $\mathbf{V}_i$  and  $\mathbf{V}_j$  are perpendicular (i.e., orthogonal) to one another. Thus vectors whose scalar product is zero are physically orthogonal to one another and, in correspondence, functions whose integrated product, as in Eq. (1.22-1) is zero are also orthogonal to one another.

Now consider that we have some arbitrary function  $f(x)$  and that we are interested in  $f(x)$  only in the range from  $x_1$  to  $x_2$ , i.e., in the interval over which the set of functions  $g(x)$  are orthogonal. Suppose further that we undertake to write  $f(x)$  as a linear sum of the functions  $g_n(x)$ . That is, we write

$$f(x) = C_1 g_1(x) + C_2 g_2(x) + \dots + C_n g_n(x) + \dots \quad (1.22-3)$$

in which the  $C$ 's are numerical coefficients. Assuming that such an expansion is indeed possible, the orthogonality of the  $g$ 's makes it very easy to compute the coefficients  $C_n$ . Thus to evaluate  $C_n$  we multiply both sides of Eq. (1.22-3) by  $g_n(x)$  and integrate over the interval of orthogonality. We have

$$\begin{aligned} \int_{x_1}^{x_2} f(x) g_n(x) dx &= C_1 \int_{x_1}^{x_2} g_1(x) g_n(x) dx \\ &+ C_2 \int_{x_1}^{x_2} g_2(x) g_n(x) dx + \dots + C_n \int_{x_1}^{x_2} g_n^2(x) dx + \dots \end{aligned} \quad (1.22-4)$$

Because of the orthogonality, all of the terms on the right-hand side of Eq. (1.22-4) become zero with a single exception and we are left with

$$\int_{x_1}^{x_2} f(x) g_n(x) dx = C_n \int_{x_1}^{x_2} g_n^2(x) dx \quad (1.22-5)$$

so that the coefficient that we are evaluating becomes

$$C_n = \frac{\int_{x_1}^{x_2} f(x) g_n(x) dx}{\int_{x_1}^{x_2} g_n^2(x) dx} \quad (1.22-6)$$

The mechanism by which we use the orthogonality of the functions to "drain" away all the terms except the term that involves the coefficient we are evaluating is often called the "orthogonality sieve."

Next suppose that each  $g_n(x)$  is selected so that the denominator of the right-hand member of Eq. (1.22-6) (which is a numerical constant) has the value

$$\int_{x_1}^{x_2} g_n^2(x) dx = 1 \quad (1.22-7)$$

In this case

$$C_n = \int_{x_1}^{x_2} f(x) g_n(x) dx \quad (1.22-8)$$

When the orthogonal functions  $g_n(x)$  are selected as in (1.22-7) they are described as being *normalized*. The use of normalized functions has the merit that the  $C_n$ 's can then be calculated from Eq. (1.22-8) and thereby avoids the need to evaluate  $\int_{x_1}^{x_2} g_n^2(x) dx$  in each case as called for in Eq. (1.22-6). A set of functions which is both orthogonal and normalized is called an *orthonormal* set.

## 1.23 COMPLETENESS OF AN ORTHOGONAL SET: THE FOURIER SERIES

Suppose on the one hand we expand a function  $f(x)$  in terms of orthogonal functions as

$$f(x) = C_1 s_1(x) + C_2 s_2(x) + C_3 s_3(x) + \dots \quad (1.23-1)$$

and on the other hand, we expand it as

$$f(x) = C_1 s_1(x) + C_3 s_3(x) + \dots \quad (1.23-2)$$

That is, in the second case, we have deliberately omitted one term. A moment's review of the procedure, described in the previous section, for evaluating coefficients makes it apparent that all the coefficients  $C_1, C_3$ , etc., that appear in both expansions will turn out to be *the same*. Hence if one expansion is correct the other is in error. We might be suspicious of the expansion of Eq. (1.23-2) on the

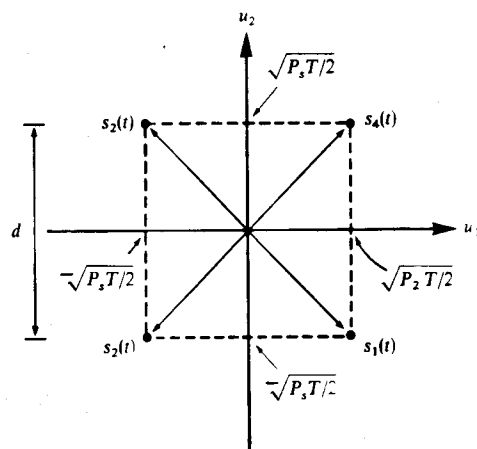


Figure 1.26-2 Signal vectors representing the four signals of Eq. (1.26-2).

It is important to note that the magnitude of the signal vectors shown in Fig. 1.26-2 is equal to the square root of the signal energy. Thus, in Eq. (1.26-2) each signal  $s_i(t)$  has a power  $P_s$  and a duration  $T$ . Thus the signal energy is  $P_s T$ , which is equal to the magnitude squared of the signal vectors as shown in Fig. 1.26-2. This result simplifies the drawing of the signal vectors in signal space.

For example, consider the signal

$$s_i(t) = \sqrt{2P_s} \cos \left[ \omega_0 t + (2i - 1) \frac{\pi}{8} \right] \quad \begin{cases} i = 1, 2, \dots, 8 \\ 0 \leq t \leq T \end{cases} \quad (1.26-6)$$

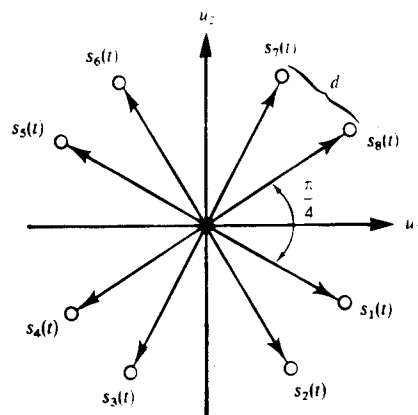


Figure 1.26-3 Signal vectors representing the four signals of Eq. (1.26-6).

Then each signal has the power  $P_s$  and duration  $T$ . Thus the magnitude of each signal vector is  $\sqrt{P_s T}$ . Furthermore, each vector is rotated  $\pi/4$  radians from the previous vector as shown in Fig. 1.26-3. The reader should verify (Prob. 1.26-1) that Fig. 1.26-3 is correct by finding the two orthonormal components  $u_1(t)$  and  $u_2(t)$  as in Eqs. (1.26-3) and (1.26-4), expanding Eq. (1.26-6) in a form similar to Eq. (1.26-5) and then plotting the result.

## REFERENCES

1. Javid, M., and E. Brenner: "Analysis of Electric Circuits," 2d ed., McGraw-Hill Book Company, New York, 1967.
2. Churchill, R. V.: "Fourier Series," McGraw-Hill Book Company, New York, 1941.
3. Papoulis, A.: "Probability, Random Variables, and Stochastic Processes," McGraw-Hill Book Company, New York, 1965.

## PROBLEMS

- 1.1-1. Verify the relationship of  $C_n$  and  $\phi_n$  to  $A_n$  and  $B_n$  as given by Eq. (1.1-6).
- 1.1-2. Calculate  $A_n$ ,  $B_n$ ,  $C_n$ , and  $\phi_n$  for a waveform  $v(t)$  which is a symmetrical square wave and which makes peak excursions to  $+\frac{1}{2}$  volt and  $-\frac{1}{2}$  volt, and has a period  $T = 1$  sec. A positive going transition occurs at  $t = 0$ .
- 1.2-1. Verify the relationship of the complex number  $V_n$  to  $C_n$  and  $\phi_n$  as given by Eq. (1.2-3).
- 1.2-2. The function

$$p(t) = \begin{cases} e^{-t} & 0 \leq t \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

is repeated every  $T = 1$  sec. Thus, with  $u(t)$  the unit step function

$$v(t) = \sum_{n=-\infty}^{\infty} p(t - n)u(t - n)$$

Find  $V_n$  and the exponential Fourier series for  $v(t)$ .

- 1.3-1. The impulse function can be defined as:

$$\delta(t) = \lim_{a \rightarrow \infty} \frac{a}{\pi} e^{-a|t|}$$

Discuss.

- 1.3-2. In Eq. (1.3-8) we see that

$$v(t) = I \sum_{k=-\infty}^{\infty} \delta(t - kT_0) = \frac{I}{T_0} + \frac{2I}{T_0} \sum_{n=1}^{\infty} \cos \frac{2\pi n t}{T_0}$$

Set  $I = 1$ ,  $T_0 = 1$ . Show by plotting

$$v(t) = 1 + 2 \sum_{n=1}^{\infty} \cos 2\pi n t$$

that the Fourier series does approximate the train of impulses.

1.4-1.  $Sa(x) \equiv (\sin x)/x$ . Determine the maxima and minima of  $Sa(x)$  and compare your result with the approximate maxima and minima obtained by letting  $x = (2n + 1)\pi/2$ ,  $n = 1, 2, \dots$

1.4-2. A train of rectangular pulses, making excursions from zero to 1 volt, have a duration of  $2 \mu\text{s}$  and are separated by intervals of  $10 \mu\text{s}$ .

(a) Assume that the center of one pulse is located at  $t = 0$ . Write the exponential Fourier series for this pulse train and plot the spectral amplitude as a function of frequency. Include at least 10 spectral components on each side of  $f = 0$ , and draw also the envelope of these spectral amplitudes.

(b) Assume that the left edge of a pulse rather than the center is located at  $t = 0$ . Correct the Fourier expansion accordingly. Does this change affect the plot of spectral amplitudes? Why not?

1.5-1. (a) A periodic waveform  $v_i(t)$  is applied to the input of a network. The output  $v_o(t)$  of the network is  $v_o(t) = \tau[dv_i(t)/dt]$ , where  $\tau$  is a constant. What is the transfer function  $H(\omega)$  of this network?

(b) A periodic waveform  $v_i(t)$  is applied to the RC network shown, whose time constant is  $\tau \equiv RC$ . Assume that the highest frequency spectral component of  $v_i(t)$  is of a frequency  $f \ll 1/\tau$ . Show that, under these circumstances, the output  $v_o(t)$  is approximately  $v_o(t) \approx \tau[dv_i(t)/dt]$ .

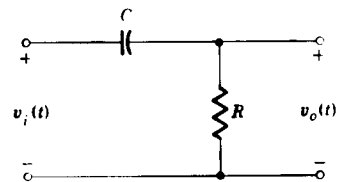


Figure P1.5-1

1.5-2. A voltage represented by an impulse train of strength  $I$  and period  $T$  is filtered by a low-pass RC filter having a 3-dB frequency  $f_c$ .

(a) Find the Fourier series of the output voltage across the capacitor.

(b) If the third harmonic of the output is to be attenuated by 1000, find  $f_c T$ .

1.6-1. Measurements on a voltage amplifier indicate a gain of 20 dB.

(a) If the input voltage is 1 volt, calculate the output voltage.

(b) If the input power is 1 mW, calculate the output power.

1.6-2. A voltage gain of 0.1 is produced by an attenuator.

(a) What is the gain in decibels?

(b) What is the power gain (not in decibels)?

1.7-1. A periodic triangular waveform  $v(t)$  is defined by

$$v(t) = \frac{2t}{T} \quad \text{for} \quad -\frac{T}{2} < t < \frac{T}{2}$$

and

$$v(t \pm T) = v(t)$$

and has the Fourier expansion

$$v(t) = \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin 2\pi n \frac{t}{T}$$

Calculate the fraction of the normalized power of this waveform which is contained in its first three harmonics.

1.7-2. The complex spectral amplitudes of a periodic waveform are given by

$$V_n = \frac{1}{|n|} e^{-j \arctan(n/2)} \quad n = \pm 1, \pm 2, \dots$$

Find the ratio of the normalized power in the second harmonic to the normalized power in the first harmonic.

1.8-1. Find  $G(f)$  for the following voltages:

(a) An impulse train of strength  $I$  and period  $T$ .

(b) A pulse train of amplitude  $A$ , duration  $\tau = I/A$ , and period  $T$ .

1.8-2. Plot  $G(f)$  for a voltage source represented by an impulse train of strength  $I$  and period  $nT$  for  $n = 1, 2, 10$ , infinity. Comment on this limiting result.

1.9-1.  $G_i(f)$  is the power spectral density of a square-wave voltage of peak-to-peak amplitude 1 and period 1. The square wave is filtered by a low-pass RC filter with a 3-dB frequency 1. The output is taken across the capacitor.

(a) Calculate  $G_o(f)$ .

(b) Find  $G_o(f)$ .

1.9-2. (a) A symmetrical square wave of zero mean value, peak-to-peak voltage 1 volt, and period 1 sec is applied to an ideal low-pass filter. The filter has a transfer function  $|H(f)| = \frac{1}{2}$  in the frequency range  $-3.5 \leq f \leq 3.5$  Hz, and  $H(f) = 0$  elsewhere. Plot the power spectral density of the filter output.

(b) What is the normalized power of the input square wave? What is the normalized power of the filter output?

1.10-1. In Eqs. (1.2-1) and (1.2-2) write  $f_0 \equiv 1/T_0$ . Replace  $f_0$  by  $\Delta f$ , i.e.,  $\Delta f$  is the frequency interval between harmonics. Replace  $n \Delta f$  by  $f$ , i.e., as  $\Delta f \rightarrow 0$ ,  $f$  becomes a continuous variable ranging from 0 to  $\infty$  as  $n$  ranges from 0 to  $\infty$ . Show that in the limit as  $\Delta f \rightarrow 0$ , so that  $\Delta f$  may be replaced by the differential  $df$ , Eq. (1.2-1) becomes

$$v(t) = \int_{-\infty}^{\infty} V(f) e^{j2\pi f t} df$$

in which  $V(f)$  is

$$V(f) = \lim_{f_0 = \Delta f \rightarrow 0} \int_{-1/2\Delta f}^{1/2\Delta f} v(t) e^{-j2\pi f t} dt = \int_{-\infty}^{\infty} v(t) e^{-j2\pi f t} dt$$

1.10-2. Find the Fourier transform of  $\sin \omega_0 t$ . Compare with the transform of  $\cos \omega_0 t$ . Plot and compare the power spectral densities of  $\cos \omega_0 t$  and  $\sin \omega_0 t$ .

1.10-3. The waveform  $v(t)$  has the Fourier transform  $V(f)$ . Show that the waveform delayed by time  $t_d$ , i.e.,  $v(t - t_d)$  has the transform  $V(f) e^{-j2\pi f t_d}$ .

1.10-4. (a) The waveform  $v(t)$  has the Fourier transform  $V(f)$ . Show that the time derivative  $(d/dt)v(t)$  has the transform  $(j2\pi f)V(f)$ .

(b) Show that the transform of the integral of  $v(t)$  is given by

$$\mathcal{F} \left[ \int_{-\infty}^t v(\lambda) d\lambda \right] = \frac{V(f)}{j2\pi f}$$

1.12-1. Derive the convolution formula in the frequency domain. That is, let  $V_1(f) = \mathcal{F}[v_1(t)]$  and  $V_2(f) = \mathcal{F}[v_2(t)]$ . Show that if  $V(f) = \mathcal{F}[v_1(t)v_2(t)]$ , then

$$V(f) = \frac{1}{2\pi} \int_{-\infty}^{\infty} V_1(\lambda) V_2(f - \lambda) d\lambda$$

or

$$V(f) = \frac{1}{2\pi} \int_{-\infty}^{\infty} V_2(\lambda) V_1(f - \lambda) d\lambda$$

1.12-2. (a) A waveform  $v(t)$  has a Fourier transform which extends over the range from  $-f_M$  to  $+f_M$ . Show that the waveform  $v^2(t)$  has a Fourier transform which extends over the range from  $-2f_M$  to  $+2f_M$ . (Hint: Use the result of Prob. 1.12-1.)

(b) A waveform  $v(t)$  has a Fourier transform  $V(f) = 1$  in the range  $-f_M$  to  $+f_M$  and  $V(f) = 0$  elsewhere. Make a plot of the transform of  $v^2(t)$ .

1.12-3. A filter has an impulse response  $h(t)$  as shown. The input to the network is a pulse of unit amplitude extending from  $t = 0$  to  $t = 2$ . By graphical means, determine the output of the filter.

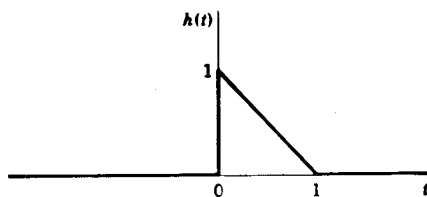


Figure P1.12-3

1.13-1. The energy of a nonperiodic waveform  $v(t)$  is

$$E = \int_{-\infty}^{\infty} v^2(t) dt$$

(a) Show that this can be written as

$$E = \int_{-\infty}^{\infty} dt v(t) \int_{-\infty}^{\infty} V(f) e^{j2\pi f t} df$$

(b) Show that by interchanging the order of integration we have

$$E = \int_{-\infty}^{\infty} V(f) V^*(f) df = \int_{-\infty}^{\infty} |V(f)|^2 df$$

which proves Eq. (1.13-5). This is an alternate proof of Parseval's theorem

1.13-2. If  $V(f) = AT \sin 2\pi fT/2\pi fT$ , find the energy  $E$  contained in  $v(t)$ .

1.13-3. A waveform  $m(t)$  has a Fourier transform  $M(f)$  whose magnitude is as shown.

(a) Find the normalized energy content of the waveform

(b) Calculate the frequency  $f_1$  such that one-half of the normalized energy is in the frequency range  $-f_1$  to  $f_1$ .

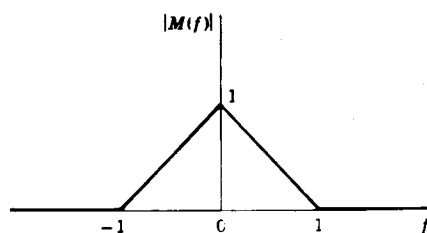


Figure P1.13-3

1.14-1. The signal  $v(t) = \cos \omega_0 t + 2 \sin 3\omega_0 t + 0.5 \sin 4\omega_0 t$  is filtered by an RC low-pass filter with a 3-dB frequency  $f_c = 2f_0$ .

(a) Find  $G_A(f)$ .

(b) Find  $G_O(f)$ .

(c) Find  $S_O$ .

1.14-2. The waveform  $v(t) = e^{-t/t_c} u(t)$  is passed through a high-pass RC circuit having a time constant equal to  $t_c$ .

(a) Find the energy spectral density at the output of the circuit.

(b) Show that the total output energy is one-half the input energy.

1.15-1. (a) An impulse of strength  $I$  is applied to a low-pass RC circuit of 3-dB frequency  $f_2$ . Calculate the output waveform.

(b) A pulse of amplitude  $A$  and duration  $\tau$  is applied to the low-pass RC circuit. Show that, if  $\tau \ll 1/f_2$ , the response at the output is approximately the response that would result from the application to the circuit of an impulse of strength  $I' = A\tau$ . Generalize this result by considering any voltage waveform of area  $I'$  and duration  $\tau$ .

1.15-2. A pulse extending from 0 to  $A$  volts and having a duration  $\tau$  is applied to a high-pass RC circuit. Show that the area under the response waveform is zero.

1.16-1. Find the cross correlation of the functions  $\sin \omega t$  and  $\cos \omega t$ .

1.16-2. Prove that  $R_{21}(\tau) = R_{12}(-\tau)$ .

1.17-1. Find the cross-correlation function  $R_{12}(\tau)$  of the two periodic waveforms shown.

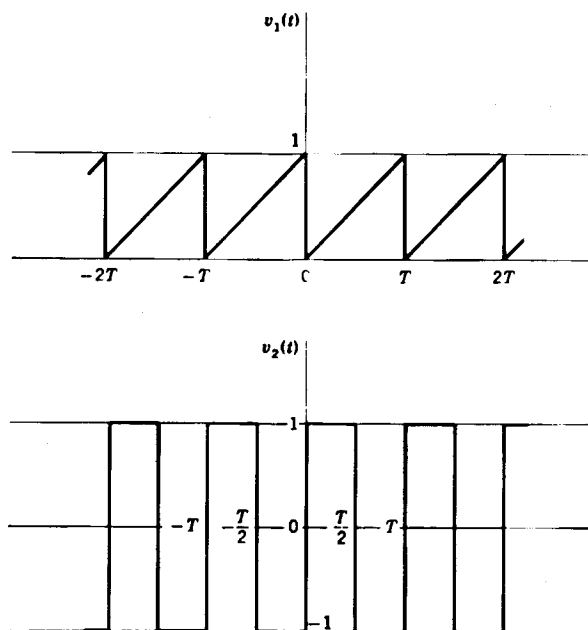


Figure P1.17-1

$$1.18-1. \quad R(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} v(t) v(t + \tau) dt$$

Prove that  $R(0) \geq R(\tau)$ . Hint: Consider

$$I = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} [v(t) - v(t + \tau)]^2 dt$$

Is  $I \geq 0$ ? Expand and integrate term by term. Show that

$$I = 2[R(0) - R(\tau)] \geq 0$$

1.18-2. Determine an expression for the correlation function of a square wave having the values 1 or 0 and a period  $T$ .

**1.19-1.** (a) Find the power spectral density of a square-wave voltage by Fourier transforming the correlation function. Use the results of Prob. 1.18-2.

(b) Compare the answer to (a) with the spectral density obtained from the Fourier series (Prob. 1.9-1a) itself.

**1.19-2.** If  $u(t) = \sin \omega_0 t$ .

(a) Find  $R(\tau)$ .

(b) If  $G(f) = \mathcal{F}[R(\tau)]$ , find  $G(f)$  directly and compare.

**1.20-1.** A waveform consists of a single pulse of amplitude  $A$  extending from  $t = -\tau/2$  to  $t = \tau/2$ .

(a) Find the autocorrelation function  $R(\tau)$  of this waveform.

(b) Calculate the energy spectral density of this pulse by evaluating  $G_E(f) = \mathcal{F}[R(\tau)]$ .

(c) Calculate  $G_E(f)$  directly by Parseval's theorem and compare.

**1.24-1.** Interchange the labels  $s_1(t)$  and  $s_2(t)$  on the waveforms of Fig. 1.24-1a and with this new labeling apply the Gram-Schmidt procedure to find expansions of the waveforms in terms of orthonormal functions.

**1.24-2.** Use the Gram-Schmidt procedure to express the functions in Fig. P1.24-2 in terms of orthonormal components.

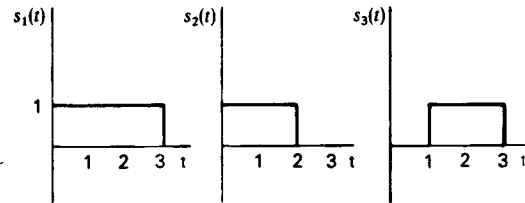


Figure P1.24-2

**1.24-3.** A set of signals ( $k = 1, 2, 3, 4$ ) is given by

$$s_k(t) = \cos\left(\omega_0 t + k \frac{\pi}{2}\right) \quad 0 \leq t \leq k \frac{2\pi}{\omega_0} = 0.$$

$$= 0 \quad \text{otherwise}$$

Use the Gram-Schmidt procedure to find an orthonormal set of functions in which the functions  $s_k(t)$  can be expanded.

**1.25-1.** Verify Eq. (1.25-10c).

**1.25-2.** (a) Show that the pythagorean theorem applies to signal functions. That is, show that if  $s_1(t)$  and  $s_2(t)$  are orthogonal then the square of the length of  $s_1(t)$  (defined as in Eq. (1.25-10b)) plus the square of the length of  $s_2(t)$  is equal to the square of the length of the sum  $s_1(t) + s_2(t)$ .

(b) Let  $s_1(t)$  and  $s_2(t)$  be the signals shown in Fig. P1.25-2. Draw the signal  $s_1(t) + s_2(t)$ . Show that  $s_1(t)$  and  $s_2(t)$  are orthogonal. Show that  $|s_1(t) + s_2(t)|^2 = |s_1(t)|^2 + |s_2(t)|^2$ .

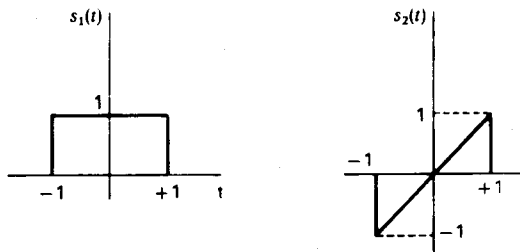


Figure P1.25-2

**1.25-3.** Verify Eq. (1.25-13).

**1.25-4.** (a) Refer to Fig. 1.25-2 and Eq. (1.25-13). Verify that the parameter  $\alpha$  in Eq. (1.25-13) is the tangent of the angle between the axes of the  $u_1, u_2$  coordinate system and the  $u'_1, u'_2$  coordinate system.

(b) Expand the functions  $s_1(t)$  and  $s_2(t)$  of Fig. 1.24-1 in terms of orthonormal functions  $u'_1(t)$  and  $u'_2(t)$  which are the axes of a coordinate system which is rotated  $60^\circ$  counterclockwise from the coordinate system whose axes are  $u_1(t)$  and  $u_2(t)$  of the same figure.

**1.26-1.** Apply the Gram-Schmidt procedure to the eight signals of Eq. (1.26-6). Verify that two orthonormal components suffice to allow a representation of any of the signals. Show that in a coordinate system of these two orthonormal signals, the geometric representation of the eight signals is as shown in Fig. 1.26-2.

**1.26-2.** A set of signals is  $s_{a1}(t) = \sqrt{2P_a} \sin \omega_0 t$ ,  $s_{a2} = -\sqrt{2P_a} \sin \omega_0 t$ . A second set of signals is  $s_{b1} = \sqrt{2P_b} \sin \omega_0 t$ ,  $s_{b2} = \sqrt{2P_b} \sin(\omega_0 t + \pi/2)$ . If the two sets of signals are to have the same distinguishability, what is the ratio  $P_b/P_a$ ?

**1.26-3.** Given two signals

$$s_1(t) = f(t) \quad 0 \leq t \leq T; \quad s_2(t) = -f(t) \quad 0 \leq t \leq T$$

Show that independently of the form of  $f(t)$  the distinguishability of the signals is given by  $\sqrt{2E}$  where  $E$  is the normalized energy of the waveform  $f(t)$ .

**1.26-4.** (a) Use the Gram-Schmidt procedure to express the functions in Fig. P1.26-4 in terms of orthonormal components. In applying the procedure, involve the functions in the order  $s_1(t)$ ,  $s_2(t)$ ,  $s_3(t)$ , and  $s_4(t)$ . Plot the functions as points in a coordinate system in which the coordinate axes are measured in units of the orthonormal functions.

(b) Repeat part (a) except that the functions are to be involved in the order  $s_1(t)$ ,  $s_4(t)$ ,  $s_3(t)$ , and  $s_2(t)$ . Plot the function:

(c) Show that the procedures of parts (a) and (b) yield the same distances between function points.

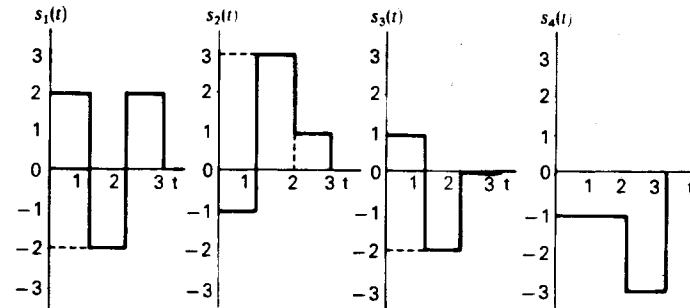


Figure P1.26-4

---

 CHAPTER  
 TWO
 

---

## RANDOM VARIABLES AND PROCESSES

A waveform which can be expressed, at least in principle, as an explicit function of time  $v(t)$  is called a *deterministic* waveform. Such a waveform is determined for all times in that, if we select an arbitrary time  $t = t_1$ , there need be no *uncertainty* about the value of  $v(t)$  at that time. The waveforms encountered in communication systems, on the other hand, are in many instances unpredictable. Consider, say, the very waveform itself which is transmitted for the purpose of communication. This waveform, which is called the *signal*, must, at least in part, be unpredictable. If such were not the case, i.e., if the signal were predictable, then its transmission would be unnecessary, and the entire communications system would serve no purpose. This point concerning the unpredictability of the signal is explored further in Chap. 13. Further, as noted briefly in Chap. 1, transmitted signals are invariably accompanied by *noise* which results from the ever-present agitation of the universe at the atomic level. These noise waveforms are also not predictable. Unpredictable waveforms such as a signal voltage  $s(t)$  or a noise voltage  $n(t)$  are examples of *random processes*. [Note that in writing symbols like  $s(t)$  and  $n(t)$  we do not imply that we can write explicit functions for these time functions.]

While random processes are not predictable, neither are they completely unpredictable. It is generally possible to predict the future performance of a random process with a certain *probability* of being correct. Accordingly, in this chapter we shall present some elemental ideas of probability theory and apply them to the description of random processes. We shall rather generally limit our discussion to the development of only those aspects of the subject which we shall have occasion to employ in this text.

2.1 PROBABILITY<sup>1</sup>

The concept of probability occurs naturally when we contemplate the possible outcomes of an experiment whose outcome is not always the same. Suppose that one of the possible outcomes is called  $A$  and that when the experiment is repeated  $N$  times the outcome  $A$  occurs  $N_A$  times. The relative frequency of occurrence of  $A$  is  $N_A/N$ , and this ratio  $N_A/N$  is not predictable unless  $N$  is very large. For example, let the experiment consist of the tossing of a die and let the outcome  $A$  correspond to the appearance of, say, the number 3 on the die. Then in 6 tosses the number 3 may not appear at all, or it may appear 6 times, or any number of times in between. Thus with  $N = 6$ ,  $N_A/N$  may be 0 or 1/6, etc., up to  $N_A/N = 1$ . On the other hand, we know from experience that when an experiment, whose outcomes are determined by chance, is repeated *very many times*, the relative frequency of a particular outcome approaches a fixed limit. Thus, if we were to toss a die very many times we would expect that  $N_A/N$  would turn out to be very close to 1/6. This limiting value of the relative frequency of occurrence is called the probability of outcome  $A$ , written  $P(A)$ , so that

$$P(A) = \lim_{N \rightarrow \infty} \frac{N_A}{N} \quad (2.1-1)$$

In many cases the experiments needed to determine the probability of an event are done more in thought than in practice. Suppose that we have 10 balls in a container, the balls being identical in every respect except that 8 are white and 2 are black. Let us ask about the probability that, in a single draw, we shall select a black ball. If we draw blindly, so that the color has no influence on the outcome, we would surely judge that the probability of drawing the black ball is 2/10. We arrive at this conclusion on the basis that we have postulated that there is absolutely nothing which favors one ball over another. There are 10 possible outcomes of the experiment, that is, any of the 10 balls may be drawn; of these 10 outcomes, 2 are favorable to our interest. The only reasonable outcome we can imagine is that, in very many drawings, 2 out of 10 will be black. Any other outcome would immediately suggest that either the black or the white balls had been favored. These considerations lead to an alternative definition of the probability of occurrence of an event  $A$ , that is

$$P(A) = \frac{\text{number of possible favorable outcomes}}{\text{total number of possible equally likely outcomes}} \quad (2.1-2)$$

It is apparent from either definition, Eq. (2.1-1) or (2.1-2), that the probability of occurrence of an event  $P$  is a positive number and that  $0 \leq P \leq 1$ . If an event is not possible, then  $P = 0$ , while if an event is certain,  $P = 1$ .

## 2.2 MUTUALLY EXCLUSIVE EVENTS

Two possible outcomes of an experiment are defined as being *mutually exclusive* if the occurrence of one outcome precludes the occurrence of the other. In this case, if the events are  $A_1$  and  $A_2$  with probabilities  $P(A_1)$  and  $P(A_2)$ , then the

the shaded area in Fig. 2.22-2 measures the probability that  $m_2$  was transmitted and read as  $m_1$ .

In Fig. 2.22-3 we represent the situation in which one of four messages  $m_1$ ,  $m_2$ ,  $m_3$ , and  $m_4$  are sent yielding receiver responses  $s_1$ ,  $s_2$ ,  $s_3$ , and  $s_4$ . The probability density function  $f_N(n)$  of the noise (multiplied by  $\frac{1}{4}$ ) is shown in Fig. 2.22-3a. Also shown are the functions  $P(m_k)f_N(r - s_k)$  for  $k = 1, 2, 3$ , and 4. We have taken all the probabilities  $P(m_k)$  to be equal at  $P(m_k) = \frac{1}{4}$ . The boundaries of the received response  $R$  at which decisions change are also shown and are marked  $\rho_{12}$ ,  $\rho_{23}$ , and  $\rho_{34}$ .

The sum of the two shaded areas shown, each of area  $\Delta$ , equals the probability  $P(E, m_2)$  that  $m_2$  is transmitted and that the message is erroneously read. Thus

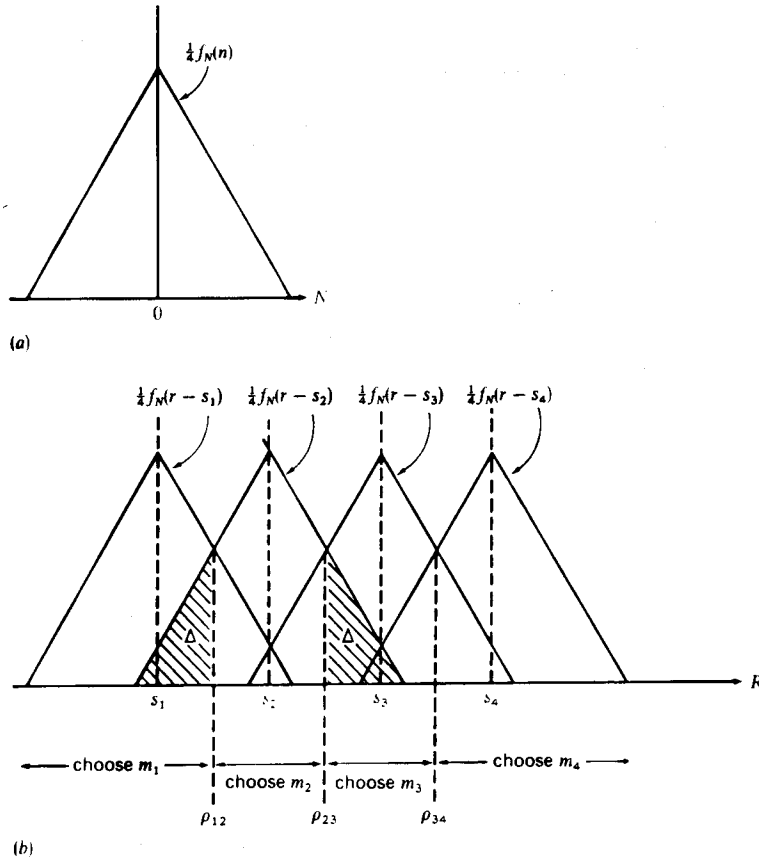


Figure 2.22-3 Probability density when four signals are transmitted. (a) Probability density of  $f_N(r)$  when noise alone is received. (b) Probability density when one of four signals is transmitted and received in noise.

$P(E, m_2) = 2\Delta$ . The remaining area under  $\frac{1}{4}f_N(r - s_2)$  is the probability that  $m_2$  is transmitted and is correctly read. This probability is  $P(C, m_2) = \frac{1}{4} - 2\Delta$ . We find of course that  $P(E, m_3) = P(E, m_2)$ . But it is most interesting to note that when we evaluate  $P(E, m_1)$  or  $P(E, m_4)$  there is only a single area  $\Delta$  involved. Hence  $P(E, m_1) = P(E, m_4) = \frac{1}{2}P(E, m_2) = \frac{1}{2}P(E, m_3)$ . Thus, we arrive at the most interesting result that given four equally likely messages, when we make boundary decisions which assure minimum average likelihood of error, the probability of making an error is not the same for all messages. Of course, this result applies for any number of equally likely messages greater than two.

The probability that a message is read correctly is

$$\begin{aligned} P(C) &= P(C, m_1) + P(C, m_2) + P(C, m_3) + P(C, m_4) \\ &= (\frac{1}{4} - \Delta) + (\frac{1}{4} - 2\Delta) + (\frac{1}{4} - 2\Delta) + (\frac{1}{4} - \Delta) \\ &= 1 - 6\Delta \end{aligned} \quad (2.22-4)$$

Thus the probability of an error is

$$P(E) = 1 - P(C) = 6\Delta \quad (2.22-5)$$

## 2.23 RANDOM PROCESSES

To determine the probabilities of the various possible outcomes of an experiment, it is necessary to repeat the experiment many times. Suppose then that we are interested in establishing the statistics associated with the tossing of a die. We might proceed in either of two ways. On one hand, we might use a single die and toss it repeatedly. Alternatively, we might toss simultaneously a very large number of dice. Intuitively, we would expect that both methods would give the same results. Thus, we would expect that a single die would yield a particular outcome, on the average, of 1 time out of 6. Similarly, with many dice we would expect that 1/6 of the dice tossed would yield a particular outcome.

Analogously, let us consider a random process such as a noise waveform  $n(t)$  mentioned at the beginning of this chapter. To determine the statistics of the noise, we might make repeated measurements of the noise voltage output of a single noise source, or we might, at least conceptually, make simultaneous measurements of the output of a very large collection of statistically identical noise sources. Such a collection of sources is called an *ensemble*, and the individual noise waveforms are called *sample functions*. A statistical average may be determined from measurements made at some fixed time  $t = t_1$  on all the sample functions of the ensemble. Thus to determine, say,  $n^2(t)$ , we would, at  $t = t_1$ , measure the voltages  $n(t_1)$  of each noise source, square and add the voltages, and divide by the (large) number of sources in the ensemble. The average so determined is the *ensemble average* of  $n^2(t_1)$ .

Now  $n(t_1)$  is a random variable and will have associated with it a probability density function. The ensemble averages will be identical with the statistical averages computed earlier in Secs. 2.11 and 2.12 and may be represented by the same

symbols. Thus the statistical or ensemble average of  $n^2(t_1)$  may be written  $E[n^2(t_1)] = \overline{n^2(t_1)}$ . The averages determined by measurements on a single sample function at successive times will yield a *time average*, which we represent as  $\langle n^2(t) \rangle$ .

In general, ensemble averages and time averages are not the same. Suppose, for example, that the statistical characteristics of the sample functions in the ensemble were changing with time. Such a variation could not be reflected in measurements made at a fixed time, and the ensemble averages would be different at different times. When the statistical characteristics of the sample functions do not change with time, the random process is described as being *stationary*. However, even the property of being stationary does not ensure that ensemble and time averages are the same. For it may happen that while each sample function is stationary the individual sample functions may differ statistically from one another. In this case, the time average will depend on the particular sample function which is used to form the average. When the nature of a random process is such that ensemble and time averages are identical, the process is referred to as *ergodic*. An ergodic process is stationary, but, of course, a stationary process is not necessarily ergodic.

Throughout this text we shall assume that the random processes with which we shall have occasion to deal are ergodic. Hence the ensemble average  $E\{n(t)\}$  is the same as the time average  $\langle n(t) \rangle$ , the ensemble average  $E\{n^2(t)\}$  is the same as the time average  $\langle n^2(t) \rangle$ , etc.

**Example 2.23-1** Consider the random process

$$V(t) = \cos(\omega_0 t + \Theta) \quad (2.23-1)$$

where  $\Theta$  is a random variable with a probability density

$$f(\theta) = \frac{1}{2\pi} \quad -\pi \leq \theta \leq \pi$$

$$= 0 \quad \text{elsewhere} \quad (2.23-2)$$

- (a) Show that the first and second moments of  $V(t)$  are independent of time.  
 (b) If the random variable  $\Theta$  in Eq. (2.23-1) is replaced by a fixed angle  $\theta_0$ , will the ensemble mean of  $V(t)$  be time independent?

**SOLUTION** (a) Choose a fixed time  $t = t_1$ . Then

$$E\{V(t_1)\} = \int_{-\pi}^{\pi} \frac{1}{2\pi} \cos(\omega_0 t_1 + \theta) d\theta = 0 \quad (2.23-3)$$

and 
$$E\{V^2(t_1)\} = \int_{-\pi}^{\pi} \frac{1}{2\pi} \cos^2(\omega_0 t_1 + \theta) d\theta = \frac{1}{2} \quad (2.23-4)$$

We note that these moments are independent of  $t_1$  and hence independent of time.

In a similar manner it can be established that all of the moments and all other statistical characteristics of  $V(t)$  are independent of time. Hence  $V(t)$  is a *stationary process*.

(b) Since  $\theta$  is known,  $V(t)$  is deterministic. For example, if  $\theta = 30^\circ$

$$E\{V(t) = V(t) = \cos(\omega_0 t + 30^\circ) \neq \text{constant} \quad (2.23-5)$$

Thus,  $V(t)$  is not stationary.

**Example 2.23-2** A voltage  $V(t)$ , which is a gaussian ergodic random process with a mean of zero and a variance of 4 volt<sup>2</sup>, is measured by a dc meter, a true rms meter, and a meter which first squares  $V(t)$  and then reads its dc component.

Find the output of each meter.

**SOLUTION** (a) The dc meter reads

$$\langle V(t) \rangle = E\{V(t)\}$$

since  $V(t)$  is ergodic. Since  $E\{V(t)\} = 0$ , the dc meter reads zero.

(b) The true rms meter reads

$$\sqrt{\langle V^2(t) \rangle} = \sqrt{E\{V^2(t)\}}$$

since  $V(t)$  is ergodic. Since  $V(t)$  has a zero mean, the true rms meter reads  $\sigma = 2$  volts.

(c) The square and average meter (a full-wave rectifier meter) yields a deflection proportional to

$$\langle v^2(t) \rangle = E\{V^2(t)\} = \sigma^2 = 4$$

## 2.24 AUTOCORRELATION

A random process  $n(t)$ , being neither periodic nor of finite energy has an autocorrelation function defined by Eq. (1.18-1). Thus

$$R(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} n(t)n(t+\tau) dt \quad (2.24-1)$$

In connection with deterministic waveforms we were able to give a physical significance to the concept of a power spectral density  $G(f)$  and to show that  $G(f)$  and  $R(\tau)$  constitute a Fourier transform pair. As an extension of that result we shall *define* the power spectral density of a random process in the same way. Thus for a random process we take  $G(f)$  to be

$$G(f) = \mathcal{F}[R(\tau)] = \int_{-\infty}^{\infty} R(\tau) e^{-j\omega\tau} d\tau \quad (2.24-2)$$



It is of interest to inquire whether  $G(f)$  defined in Eq. (2.24-2) for a random process has a physical significance which corresponds to the physical significance of  $G(f)$  for deterministic waveforms.

For this purpose consider a *deterministic* waveform  $v(t)$  which extends from  $-\infty$  to  $\infty$ . Let us select a section of this waveform which extends from  $-T/2$  to  $T/2$ . This waveform  $v_T(t) = v(t)$  in this range, and otherwise  $v_T(t) = 0$ . The waveform  $v_T(t)$  has a Fourier transform  $V_T(f)$ . We recall that  $|V_T(f)|^2$  is the energy spectral density; that is,  $|V_T(f)|^2 df$  is the normalized energy in the spectral range  $df$ . Hence, over the interval  $T$  the normalized power density is  $|V_T(f)|^2/T$ . As  $T \rightarrow \infty$ ,  $v_T(t) \rightarrow v(t)$ , and we then have the result that the physical significance of the power spectral density  $G(f)$ , at least for a deterministic waveform, is that

$$G(f) = \lim_{T \rightarrow \infty} \frac{1}{T} |V_T(f)|^2 \quad (2.24-3)$$

Correspondingly, we state, without proof, that when  $G(f)$  is defined for a random process, as in Eq. (2.24-2), as the transform of  $R(\tau)$ , then  $G(f)$  has the significance that

$$G(f) = \lim_{T \rightarrow \infty} E \left\{ \frac{1}{T} |N_T(f)|^2 \right\} \quad (2.24-4)$$

where  $E\{\}$  represents the ensemble average or expectation and  $N_T(f)$  represents the Fourier transform of a truncated section of a sample function of the random process  $n(t)$ .

The autocorrelation function  $R(\tau)$  is, as indicated in Eq. (2.24-1), a *time average* of the product  $n(t)$  and  $n(t + \tau)$ . Since we have assumed an ergodic process, we are at liberty to perform the averaging over any sample function of the ensemble, since every sample function will yield the same result. However, again because the noise process is ergodic, we may replace the time average by an ensemble average and write, instead of Eq. (2.24-1),

$$R(\tau) = E\{n(t)n(t + \tau)\} \quad (2.24-5)$$

The averaging indicated in Eq. (2.24-5) has the following significance: At some *fixed* time  $t$ ,  $n(t)$  is a random variable, the possible values for which are the values  $n(t)$  assumed at time  $t$  by the individual sample functions of the ensemble. Similarly, at the *fixed* time  $t + \tau$ ,  $n(t + \tau)$  is also a random variable. It then appears that  $R(\tau)$  as expressed in Eq. (2.24-5) is the covariance between these two random variables.

Suppose then that we should find that for some  $\tau$ ,  $R(\tau) = 0$ . Then the random variables  $n(t)$  and  $n(t + \tau)$  are uncorrelated, and for the gaussian process of interest to us,  $n(t)$  and  $n(t + \tau)$  are independent. Hence, if we should select some sample function, a knowledge of the value of  $n(t)$  at time  $t$  would be of no assistance in improving our ability to predict the value attained by that same sample function at time  $t + \tau$ .

The physical fact about the noise, which is of principal concern in connection with communications systems, is that such noise has a power spectral density

$G(f)$  which is uniform over all frequencies. Such noise is referred to as “white” noise in analogy with the consideration that white light is a combination of all colors, that is, colors of all frequencies. Actually as is pointed out in Sec. 14.5, there is an upper-frequency limit beyond which the spectral density falls off sharply. However, this upper-frequency limit is so high that we may ignore it for our purposes.

Now, since the autocorrelation  $R(\tau)$  and the power spectral density  $G(f)$  are a Fourier transform pair, they have the properties of such pairs. Thus when  $G(f)$  extends over a wide frequency range,  $R(\tau)$  is restricted to a narrow range of  $\tau$ . In the limit, if  $G(f) = I$  (a constant) for all frequencies from  $-\infty \leq f \leq +\infty$ , then  $R(\tau)$  becomes  $R(\tau) = I \delta(\tau)$ , where  $\delta(\tau)$  is the delta function with  $\delta(\tau) = 0$  except for  $\tau = 0$ . Since, then, for white noise,  $R(\tau) = 0$  except for  $\tau = 0$ , Eq. (2.24-5) says that  $n(t)$  and  $n(t + \tau)$  are uncorrelated and hence independent, no matter how small  $\tau$ .

## 2.25 POWER SPECTRAL DENSITY OF A SEQUENCE OF RANDOM PULSES

We shall occasionally need to have information about the power spectral density of a sequence of random pulses such as is indicated in Fig. 2.25-1. The pulses are of the same form but have random amplitudes and statistically independent random times of occurrence. The waveform (the random process) is stationary so that the statistical features of the waveforms are time invariant. Correspondingly, there is an invariant *average* time of separation  $T_s$  between pulses. We further assume that there is no overlap between pulses.

If the Fourier transform of a single sample pulse  $P_1(t)$  is  $P_1(f)$  then Parseval's theorem [Eq. (1.13-5)] states that the normalized energy of the pulse is

$$E_1 = \int_{-\infty}^{\infty} P_1(f) P_1^*(f) df = \int_{-\infty}^{\infty} |P_1(f)|^2 df \quad (2.25-1)$$

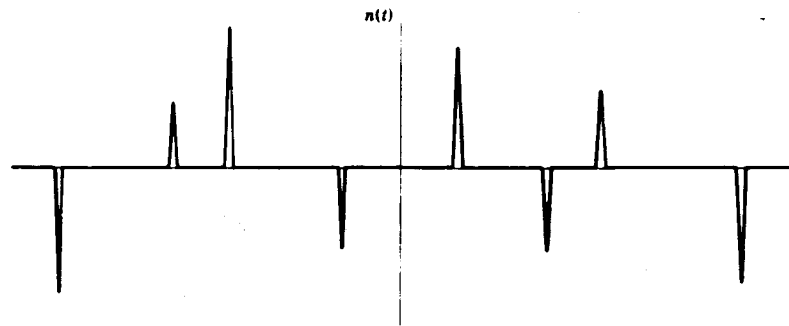


Figure 2.25-1 Pulses of random amplitude and time of occurrence.

## PROBLEMS

- 2.1-1.** Six dice are thrown simultaneously. What is the probability that at least 1 die shows a 3?
- 2.1-2.** A card is drawn from a deck of 52 cards.
- What is the probability that a 2 is drawn?
  - What is the probability that a 2 of clubs is drawn?
  - What is the probability that a spade is drawn?
- 2.2-1.** A card is picked from each of four 52-card decks of cards.
- What is the probability of selecting at least one 6 of spades?
  - What is the probability of selecting at least 1 card larger than an 8?
- 2.3-1.** A card is drawn from a 52-card deck, and without replacing the first card a second card is drawn. The first and second cards are not replaced and a third card is drawn.
- If the first card is a heart, what is the probability of the second card being a heart?
  - If the first and second cards are hearts, what is the probability that the third card is the king of clubs?
- 2.3-2.** Two factories produce identical clocks. The production of the first factory consists of 10,000 clocks of which 100 are defective. The second factory produces 20,000 clocks of which 300 are defective. What is the probability that a particular defective clock was produced in the first factory?
- 2.3-3.** One box contains two black balls. A second box contains one black and one white ball. We are told that a ball was withdrawn from one of the boxes and that it turned out to be black. What is the probability that this withdrawal was made from the box that held the two black balls?
- 2.4-1.** Two dice are tossed.
- Find the probability of a 3 and a 4 appearing
  - Find the probability of a 7 being rolled.
- 2.4-2.** A card is drawn from a deck of 52 cards, then replaced, and a second card drawn.
- What is the probability of drawing the same card twice?
  - What is the probability of first drawing a 3 of hearts and then drawing a 4 of spades?
- 2.6-1.** A die is tossed and the number appearing is  $n_i$ . Let  $N$  be the random variable identifying the outcome of the toss defined by the specifications  $N = n_i$  when  $n_i$  appears. Make a plot of the probability  $P(N \leq n_i)$  as a function of  $n_i$ .
- 2.6-2.** A coin is tossed four times. Let  $H$  be the random variable which identifies the number of heads which occur in these four tosses. It is defined by  $H = h$ , where  $h$  is the number of heads which appear. Make a plot of the probability  $P(H \leq h)$  as a function of  $h$ .
- 2.6-3.** A coin is tossed until a head appears. Let  $T$  be the random variable which identifies the number of tosses  $t$  required for the appearance of this first head. Make a plot of the probability  $P(T \leq t)$  as a function of  $t$  up to  $t = 5$ .

**2.7-1.** An important probability density function is the Rayleigh density

$$f(x) = \begin{cases} xe^{-x^2/2} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

- Prove that  $f(x)$  satisfies Eqs. (2.7-2) and (2.7-3).
  - Find the distribution function  $F(x)$ .
- 2.8-1.** Refer to Fig. 2.6-1.
- Find  $P(2 < n \leq 11)$
  - Find  $P(2 \leq n < 11)$ .
  - Find  $P(2 \leq n \leq 11)$ .
  - Find  $F(9)$ .
- 2.8-2.** Refer to the Rayleigh density function given in Prob. 2.7-1. Find the probability  $P(x_1 < x \leq x_2)$ , where  $x_2 - x_1 = 1$ , so that  $P(x_1 < x \leq x_2)$  is a maximum. Hint: Find  $P(x_1 < x \leq x_2)$ , replace  $x_2$  by  $1 + x_1$ , and maximize  $P$  with respect to  $x_1$ .

**2.8-3.** Refer to the Rayleigh density function given in Prob. 2.7-1. Find

- $P(0.5 < x \leq 2)$
- $P(0.5 \leq x < 2)$ .

**2.9-1.** The joint probability density of the random variables  $X$  and  $Y$  is  $f(x, y) = ke^{-(x+y)}$  in the range  $0 \leq x \leq \infty, 0 \leq y \leq \infty$ , and  $f(x, y) = 0$  otherwise.

- Find the value of the constant  $k$
- Find the probability density  $f(x)$ , the probability density of  $X$  independently of  $Y$ .
- Find the probability  $P(0 \leq X \leq 2; 2 \leq Y \leq 3)$ .
- Are the random variables dependent or independent?

**2.9-2.**  $X$  is a random variable having a gaussian density.  $E(X) = 0, \sigma_x^2 = 1$ .  $V$  is a random variable having the values 1 or  $-1$ , each with probability  $\frac{1}{2}$ .

- Find the joint density  $f_{X,V}(x, v)$ .
- Show that  $f_V(v) = \int_{-\infty}^{\infty} f_{X,V}(x, v) dx$ .

**2.9-3.** The joint probability density of the random variables  $X$  and  $Y$  is  $f(x, y) = xe^{-x(y+1)}$  in the range  $0 \leq x \leq \infty, 0 \leq y \leq \infty$ , and  $f(x, y) = 0$  otherwise.

- Find  $f(x)$  and  $f(y)$ , the probability density of  $X$  independently of  $Y$  and  $Y$  independently of  $X$ .

- Are the random variables dependent or independent?

**2.10-1.** (a) In a communication channel as represented in Fig. 2.10-1  $P(r_0|m_0) = 0.9, P(r_1|m_0) = 0.1, P(r_0|m_1) = 0.4$ , and  $P(r_1|m_1) = 0.6$ . On a single set of coordinate axes make plots of  $P(m_0|r_0), P(m_1|r_0), P(m_0|r_1)$  and  $P(m_1|r_1)$  as a function of  $P(m_0)$ . Mark the range of  $m_0$  for which the algorithm of Eq. (2.10-1) prescribes that we choose  $m_0$  if  $r_0$  is received and  $m_1$  if  $r_1$  is received, the range for which we choose  $m_0$  no matter what is received and the range for which we choose  $m_1$  no matter what is received.

- Calculate and plot the probability of error as a function of  $P(m_0)$ .

**2.10-2.** (a) For the channel and message probabilities given in Fig. 2.10-2 determine the best decisions about the transmitted message for each possible received response.

- With decisions made as in part (a) calculate the probability of error.

(c) Suppose the decision-making apparatus at the receiver were inoperative so that at the receiver nothing could be determined except that a message had been received. What would be the best strategy for determining what message had been transmitted and what would be the corresponding error probability?

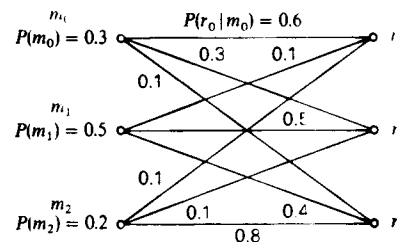


Figure P2.10-2

**2.11-1.** If  $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$  for all  $x$ , show that

- $E(X^{2n}) = 1 \cdot 3 \cdot 5 \cdots (n-1), n = 1, 2, \dots$
- $E(X^{2n-1}) = 0, n = 1, 2, \dots$

**2.11-2.** Compare the most probable [ $f(x)$  is a maximum] and the average value of  $X$  when

- $f_{X_1}(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-m)^2/2}$  for all  $x$
- $f_{X_2}(x) = \begin{cases} xe^{-x^2/2} & \text{for } x \geq 0 \\ 0 & \text{elsewhere} \end{cases}$

2.12-1. Calculate the variance of the random variables having densities:

- (a) The gaussian density  $f_{X_1}(x) = \frac{1}{\sqrt{2\tau}} e^{-(x-m)^2/2\tau}$ , all  $x$ .  
 (b) The Rayleigh density  $f_{X_2}(x) = xe^{-x^2/2}$ ,  $x \geq 0$ .  
 (c) The uniform density  $f_{X_3}(x) = 1/a$ ,  $-a/2 \leq x \leq a/2$ .

2.12-2. Consider the Cauchy density function

$$f(x) = \frac{K}{1+x^2} \quad -\infty \leq x \leq \infty$$

- (a) Find  $K$  so that  $f(x)$  is a density function.  
 (b) Find  $E(X)$ .  
 (c) Find the variance of  $X$ . Comment on the significance of this result.

2.12-3. The random variable  $X$  has a variance  $\sigma^2$  and a mean  $m$ . The random variable  $Y$  is related to  $X$  by  $Y = aX + b$ , where  $a$  and  $b$  are constants. Find the mean and variance of  $Y$ .

2.13-1. A probability density function  $f(x)$  is uniform over the range from  $-L$  to  $+L$ .

- (a) Calculate and plot the probability  $P[|x| \leq \epsilon]$  as a function of  $\epsilon$ .  
 (b) Calculate and plot the quantity  $\sigma^2/\epsilon^2$  and verify that Tchebycheff's rule is valid in this case.

2.14-1. Refer to the gaussian density given in Eq. (2.14-1).

- (a) Show that  $E[(X-m)^{2n-1}] = 0$ .  
 (b) Show that  $E[(X-m)^{2n}] = 1 \cdot 3 \cdot 5 \cdots (n-1)\sigma^2$ .

2.14-2. Given a gaussian probability function  $f(x)$  of mean value zero and variance  $\sigma^2$ .

- (a) As a function of  $\sigma$ , plot the probability  $P[|x| \geq \epsilon]$ .  
 (b) Calculate and plot the quantity  $\sigma^2/\epsilon^2$  and verify that Tchebycheff's rule is valid in this case.

2.14-3. A random variable  $V = b + X$ , where  $X$  is a gaussian distributed random variable with mean 0 and variance  $\sigma^2$ , and  $b$  is a constant. Show that  $V$  is a gaussian distributed random variable with mean  $b$  and variance  $\sigma^2$ .

2.14-4. The joint density function of two dependent variables  $X$  and  $Y$  is

$$f(x, y) = \frac{1}{\pi\sqrt{1-\rho^2}} e^{-2(x^2 - \rho xy + y^2)/(1-\rho^2)}$$

- (a) Show that, when  $X$  and  $Y$  are each considered without reference to the other, each is a gaussian variable, i.e.,  $f(x)$  and  $f(y)$  are gaussian density functions.  
 (b) Find  $\sigma_x^2$  and  $\sigma_y^2$ .

2.15-1. Obtain values for and plot  $\text{erf } u$  versus  $u$ .

2.15-2. On the same set of axes used in Prob. 2.15-1 plot  $e^{-u^2}$  and  $\text{erfc } u$  versus  $u$ . Compare your results.

2.15-3. The probability

$$P_{\pm k\sigma} \equiv P(m - k\sigma \leq X \leq m + k\sigma) = \int_{m-k\sigma}^{m+k\sigma} \frac{e^{-(x-m)^2/2\sigma^2}}{\sqrt{2\pi}\sigma} dx$$

- (a) Change variables by letting  $u = (x-m)/\sqrt{2}\sigma$ .  
 (b) Show that  $P_{\pm k\sigma} = \text{erf}(k/\sqrt{2})$ .

2.16-1. Show that a random variable with a Rayleigh density as in Eq. (2.16-1) has a mean value  $R = \sqrt{(\pi/2)}\alpha$ , a mean square value  $R^2 = 2\alpha^2$ , and a variance  $\sigma^2 = (2 - \pi/2)\alpha^2$ .

2.16-2. (a) A voltage  $V$  is a function of time  $t$  and is given by

$$V(t) = X \cos \omega t + Y \sin \omega t$$

in which  $\omega$  is a constant angular frequency and  $X$  and  $Y$  are independent gaussian variables each with zero mean and variance  $\sigma^2$ . Show that  $V(t)$  may be written

$$V(t) = R \cos(\omega t + \Theta)$$

in which  $R$  is a random variable with a Rayleigh probability density and  $\Theta$  is a random variable with uniform density.

- (b) If  $\sigma^2 = 1$ , what is the probability that  $R \geq 1$ ?

2.17-1. Derive Eq. (2.17-6) directly from the definition  $\sigma_z^2 = E\{(Z - m_z)^2\}$ .

2.17-2.  $Z = X_1 + X_2 + \cdots + X_N$ ,  $E(X_i) = m$ .

- (a) Find  $E(Z)$ .

$$(b) \text{ If } E(X_i X_j) = \begin{cases} 1 & j = i \\ \rho & j = i \pm 1 \\ 0 & \text{otherwise} \end{cases}$$

find (1)  $E(Z^2)$  and (2)  $\sigma_z^2$ .

2.18-1. The independent random variables  $X$  and  $Y$  are added to form  $Z$ . If

$$f_X(x) = xe^{-x^2/2} \quad 0 \leq x \leq \infty \quad \text{and} \quad f_Y(y) = \frac{1}{2}e^{-|y|} \quad |y| < \infty$$

find  $f_Z(z)$ .

2.18-2. The independent random variables  $X$  and  $Y$  have the probability densities

$$f(x) = e^{-x} \quad 0 \leq x \leq \infty \\ f(y) = 2e^{-2y} \quad 0 \leq y \leq \infty$$

Find and plot the probability density of the variable  $Z = X + Y$ .

2.18-3. The random variable  $X$  has a probability density uniform in the range  $0 \leq x \leq 1$  and zero elsewhere. The independent variable  $Y$  has a density uniform in the range  $0 \leq x \leq 2$  and zero elsewhere. Find and plot the density of  $Z = X + Y$ .

2.18-4. The  $N$  independent gaussian random variables  $X_1, \dots, X_N$  are added to form  $Z$ . If the mean of  $X_i$  is 1 and its variance is 1, find  $f_Z(z)$ .

2.19-1. Two gaussian random variables  $X$  and  $Y$ , each with mean zero and variance  $\sigma^2$ , between which there is a correlation coefficient  $\rho$ , have a joint probability density given by

$$f(x, y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp \left[ -\frac{x^2 - 2\rho xy + y^2}{2\sigma^2(1-\rho^2)} \right]$$

(a) Verify that the symbol  $\rho$  in the expression for  $f(x, y)$  is indeed the correlation coefficient. That is, evaluate  $E\{XY\}/\sigma^2$  and show that the result is  $\rho$  as required by Eq. (2.19-5).

- (b) Show that the case  $\rho = 0$  corresponds to the circumstance where  $X$  and  $Y$  are independent.

2.19-2. The random variables  $X$  and  $Y$  are related to the random variable  $\Theta$  by  $X = \sin \Theta$  and  $Y = \cos \Theta$ . The variable  $\Theta$  has a uniform probability density in the range from 0 to  $2\pi$ . Show that  $X$  and  $Y$  are not independent but that, nonetheless,  $E(XY) = 0$  so that they are uncorrelated.

2.19-3. The random variables  $X_1, X_2, X_3, \dots$  are dependent but uncorrelated.  $Z = X_1 + X_2 + X_3 + \dots$ . Show that  $\sigma_z^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \dots$ .

2.20-1. The random variables  $X_1, X_2, X_3$  are independent and each has a uniform probability density in the range  $0 \leq x \leq 1$ . Find and plot the probability density of  $X_1 + X_2$  and of  $X_1 + X_2 + X_3$ .

2.21-1. Verify Eq. (2.21-16).

2.21-2. In a communication system used to transmit a sequence of messages, it is known that the error probability is  $4 \times 10^{-5}$ . A sample survey of  $N$  messages is made. It is required that, with a probability not to exceed 0.05, the error rate in the sample is not to be larger than  $5 \times 10^{-5}$ . How many messages must be included in the sample?

2.22-1. A communication channel transmits, in random order, two messages  $m_1$  and  $m_2$ . The message  $m_1$  occurs three times more frequently than  $m_2$ . Message  $m_1$  generates a receiver response  $r_1 = -1V$  while  $m_2$  generates  $r_2 = +1V$ . The channel is corrupted by noise  $n$  with a uniform probability density which extends from  $n = -1.5V$  to  $n = +1.5V$ .

- (a) Find the probability that  $m_1$  is mistaken for  $m_2$  and the probability that  $m_2$  is mistaken for  $m_1$ .

- (b) What is the probability that the receiver will determine a message correctly?

**2.22-2.** A communication channel transmits, in random order, two messages  $m_1$  and  $m_2$  with equal likelihood. Message  $m_1$  generates response  $r_1 = -1V$  at the receiver and message  $m_2$  generates  $r_2 = +1V$ . The channel is corrupted with gaussian noise with variance  $\sigma^2 = 1$  volt<sup>2</sup>. Find the probability that the receiver will determine a message correctly.

**2.22-3.** A communication channel transmits, in random order, two messages  $m_1$  and  $m_2$ . The message  $m_1$  occurs three times more frequently than  $m_2$ . Message  $m_1$  generates a receiver response  $r_1 = +1V$  while  $m_2$  generates  $r_2 = -1V$ . The channel is corrupted by noise  $n$  whose probability density has the triangular form shown in Fig. 2.22-2a with  $f_N(n) = 0$  at  $f_N(n) = -2V$  and at  $f_N(n) = +2V$ .

(a) What are the ranges of  $r$  for which the decision is to be made that  $m_1$  was transmitted and for which the decision is to be made that  $m_2$  was transmitted?

(b) What is the probability that the message will be read correctly?

**2.23-1.** The function of time  $Z(t) = X_1 \cos \omega_0 t - X_2 \sin \omega_0 t$  is a random process. If  $X_1$  and  $X_2$  are independent gaussian random variables each with zero mean and variance  $\sigma^2$ , find

(a)  $E(Z)$ ,  $E(Z^2)$ ,  $\sigma_z^2$ , and

(b)  $f_Z(z)$ .

**2.23-2.**  $Z(t) = M(t) \cos(\omega_0 t + \Theta)$ .  $M(t)$  is a random process, with  $E(M(t)) = 0$  and  $E(M^2(t)) = M_0^2$ .

(a) If  $\Theta = 0$ , find  $E(Z^2)$ . Is  $Z(t)$  stationary?

(b) If  $\Theta$  is an independent random variable such that  $f_\Theta(\theta) = 1/2\pi$ ,  $-\pi \leq \theta \leq \pi$ , show that  $E(Z^2(t)) = E(M^2(t))E(\cos^2(\omega_0 t + \Theta)) = M_0^2/2$ . Is  $Z(t)$  now stationary?

**2.24-1.** Refer to Prob. 2.23-1. Find  $R_z(\tau)$ .

**2.24-2.** A random process  $n(t)$  has a power spectral density  $G(f) = \eta/2$  for  $-\infty \leq f \leq \infty$ . The random process is passed through a low-pass filter which has a transfer function  $H(f) = 2$  for  $-f_M \leq f \leq f_M$  and  $H(f) = 0$  otherwise. Find the power spectral density of the waveform at the output of the filter.

**2.24-3.** White noise  $n(t)$  with  $G(f) = \eta/2$  is passed through a low-pass RC network with a 3-dB frequency  $f_c$ .

(a) Find the autocorrelation  $R(\tau)$  of the output noise of the network.

(b) Sketch  $\rho(\tau) = R(\tau)/R(0)$ .

(c) Find  $\omega_c \tau$  such that  $\rho(\tau) \leq 0.1$ .

**2.25-1.** Consider a train of rectangular pulses. The  $k$ th pulse has a width  $\tau$  and a height  $A_k$ .  $A_k$  is a random variable which can have the values 1, 2, 3, ..., 10 with equal probability. Assuming statistical independence between amplitudes and assuming that the average separation between pulses is  $T_s$ , find the power spectral density  $G_n(f)$  of the pulse train.

**2.25-2.** A pulse train consists of rectangular pulses having an amplitude of 2 volts and widths which are either 1  $\mu$ s or 2  $\mu$ s with equal probability. The mean time between pulses is 5  $\mu$ s. Find the power spectral density  $G_n(f)$  of the pulse train.

**2.26-1.** Consider the power spectral density of an NRZ waveform as given by Eq. (2.26-4) and as shown in Fig. 2.26-3. By consulting a table of integrals, show that the power of the NRZ waveform is reduced by only 10 percent if the waveform is passed through an ideal low-pass filter with cutoff at  $f = f_b$ .

**2.26-2.** Consider the power spectral density of a biphasic waveform as given by Eq. (2.26-7) and as shown in Fig. 2.26-4. By consulting a table of integrals, show that, if the waveform is passed through an ideal low-pass filter with cutoff at  $2f_b$ , 95 percent of the power will be passed. Show also that, if the filter cutoff is set at  $f_b$ , then only 70 percent of the power is transmitted.

**2.27-1.** An NRZ waveform consists of alternating 0's and 1's. The bit interval is 1  $\mu$ s and the waveform makes excursions between +1V and -1V. The waveform is transmitted through an RC high-pass filter of time constant 1  $\mu$ s. Draw the output waveform and calculate numerical values of all details of the waveform.

**2.27-2.** An NRZ waveform consists of alternating 0's and 1's. The bit interval is 1  $\mu$ s and the waveform makes excursions between +1V and -1V. The waveform is transmitted through an RC low-pass filter of time constant 1  $\mu$ s. Draw the output waveform and calculate numerical values of all details of the waveform.

## CHAPTER THREE

### AMPLITUDE-MODULATION SYSTEMS

One of the basic problems of communication engineering is the design and analysis of systems which allow many individual messages to be transmitted simultaneously over a single communication channel. A method by which such multiple transmission, called *multiplexing*, may be achieved consists in translating each message to a different position in the *frequency* spectrum. Such multiplexing is called *frequency multiplexing*. The individual message can eventually be separated by filtering. Frequency multiplexing involves the use of an auxiliary waveform, usually sinusoidal, called a *carrier*. The operations performed on the signal to achieve frequency multiplexing result in the generation of a waveform which may be described as the carrier modified in that its amplitude, frequency, or phase, individually or in combination, varies with time. Such a modified carrier is called a *modulated carrier*. In some cases the modulation is related simply to the message; in other cases the relationship is quite complicated. In this chapter, we discuss the generation and characteristics of amplitude-modulated carrier waveforms.<sup>1</sup>

#### 3.1 FREQUENCY TRANSLATION

It is often advantageous and convenient, in processing a signal in a communications system, to translate the signal from one region in the frequency domain to another region. Suppose that a signal is bandlimited, or nearly so, to the frequency range extending from a frequency  $f_1$  to a frequency  $f_2$ . The process of frequency translation is one in which the original signal is replaced with a new

2-2. A communication channel transmits, in random order, two messages  $m_1$  and  $m_2$  with equal likelihood. Message  $m_1$  generates response  $r_1 = -1V$  at the receiver and message  $m_2$  generates  $r_2 = +1V$ . The channel is corrupted with gaussian noise with variance  $\sigma^2 = 1 \text{ volt}^2$ . Find the probability that the receiver will determine a message correctly.

2-3. A communication channel transmits, in random order, two messages  $m_1$  and  $m_2$ . The message occurs three times more frequently than  $m_2$ . Message  $m_1$  generates a receiver response  $r_1 = +1V$  and message  $m_2$  generates  $r_2 = -1V$ . The channel is corrupted by noise  $n$  whose probability density has the angular form shown in Fig. 2.22-2a with  $f_N(n) = 0$  at  $f_N(n) = -2V$  and at  $f_N(n) = +2V$ .

(a) What are the ranges of  $r$  for which the decision is to be made that  $m_1$  was transmitted and which the decision is to be made that  $m_2$  was transmitted?

(b) What is the probability that the message will be read correctly?

3-1. The function of time  $Z(t) = X_1 \cos \omega_0 t - X_2 \sin \omega_0 t$  is a random process. If  $X_1$  and  $X_2$  are independent gaussian random variables each with zero mean and variance  $\sigma^2$ , find

(a)  $E\{Z\}$ ,  $E\{Z^2\}$ ,  $\sigma_z^2$ , and

(b)  $f_z(f)$ .

3-2.  $Z(t) = M(t) \cos(\omega_0 t + \Theta)$ .  $M(t)$  is a random process, with  $E\{M(t)\} = 0$  and  $E\{M^2(t)\} = M_0^2$ .

(a) If  $\Theta = 0$ , find  $E\{Z^2\}$ . Is  $Z(t)$  stationary?

(b) If  $\Theta$  is an independent random variable such that  $f_\Theta(\theta) = 1/2\pi$ ,  $-\pi \leq \theta \leq \pi$ , show that  $E\{Z(t)\} = E\{M^2(t)\}E\{\cos^2(\omega_0 t + \Theta)\} = M_0^2/2$ . Is  $Z(t)$  now stationary?

3-1. Refer to Prob. 2.23-1. Find  $R_z(\tau)$ .

3-2. A random process  $n(t)$  has a power spectral density  $G(f) = \eta/2$  for  $-\infty \leq f \leq \infty$ . The random process is passed through a low-pass filter which has a transfer function  $H(f) = 2$  for  $-f_M \leq f \leq f_M$  and  $H(f) = 0$  otherwise. Find the power spectral density of the waveform at the output of the filter.

3-3. White noise  $n(t)$  with  $G(f) = \eta/2$  is passed through a low-pass RC network with a 3-dB frequency  $f_c$ .

(a) Find the autocorrelation  $R(\tau)$  of the output noise of the network.

(b) Sketch  $\rho(\tau) = R(\tau)/R(0)$ .

(c) Find  $\omega_c \tau$  such that  $\rho(\tau) \leq 0.1$ .

3-1. Consider a train of rectangular pulses. The  $k$ th pulse has a width  $\tau$  and a height  $A_k$ .  $A_k$  is a discrete variable which can have the values 1, 2, 3, ..., 10 with equal probability. Assuming statistical independence between amplitudes and assuming that the average separation between pulses is  $T_s$ , find power spectral density  $G_p(f)$  of the pulse train.

3-2. A pulse train consists of rectangular pulses having an amplitude of 2 volts and widths which are either 1  $\mu\text{s}$  or 2  $\mu\text{s}$  with equal probability. The mean time between pulses is 5  $\mu\text{s}$ . Find the power spectral density  $G_p(f)$  of the pulse train.

3-1. Consider the power spectral density of an NRZ waveform as given by Eq. (2.26-4) and as shown in Fig. 2.26-3. By consulting a table of integrals, show that the power of the NRZ waveform is reduced by only 10 percent if the waveform is passed through an ideal low-pass filter with cutoff at  $f_c$ .

3-2. Consider the power spectral density of a biphasic waveform as given by Eq. (2.26-7) and as shown in Fig. 2.26-4. By consulting a table of integrals, show that, if the waveform is passed through an ideal low-pass filter with cutoff at  $2f_c$ , 95 percent of the power will be passed. Show also that, if the cutoff is set at  $f_c$ , then only 70 percent of the power is transmitted.

3-1. An NRZ waveform consists of alternating 0's and 1's. The bit interval is 1  $\mu\text{s}$  and the waveform makes excursions between  $+1V$  and  $-1V$ . The waveform is transmitted through an RC high-pass filter of time constant 1  $\mu\text{s}$ . Draw the output waveform and calculate numerical values of all peaks of the waveform.

3-2. An NRZ waveform consists of alternating 0's and 1's. The bit interval is 1  $\mu\text{s}$  and the waveform makes excursions between  $+1V$  and  $-1V$ . The waveform is transmitted through an RC low-pass filter of time constant 1  $\mu\text{s}$ . Draw the output waveform and calculate numerical values of all peaks of the waveform.

## AMPLITUDE-MODULATION SYSTEMS

One of the basic problems of communication engineering is the design and analysis of systems which allow many individual messages to be transmitted simultaneously over a single communication channel. A method by which such multiple transmission, called *multiplexing*, may be achieved consists in translating each message to a different position in the frequency spectrum. Such multiplexing is called *frequency multiplexing*. The individual message can eventually be separated by filtering. Frequency multiplexing involves the use of an auxiliary waveform, usually sinusoidal, called a *carrier*. The operations performed on the signal to achieve frequency multiplexing result in the generation of a waveform which may be described as the carrier modified in that its amplitude, frequency, or phase, individually or in combination, varies with time. Such a modified carrier is called a *modulated carrier*. In some cases the modulation is related simply to the message; in other cases the relationship is quite complicated. In this chapter, we discuss the generation and characteristics of amplitude-modulated carrier waveforms.<sup>1</sup>

## 3.1 FREQUENCY TRANSLATION

It is often advantageous and convenient, in processing a signal in a communications system, to translate the signal from one region in the frequency domain to another region. Suppose that a signal is bandlimited, or nearly so, to the frequency range extending from a frequency  $f_1$  to a frequency  $f_2$ . The process of frequency translation is one in which the original signal is replaced with a new

signal whose spectral range extends from  $f'_1$  to  $f'_2$  and which *new* signal bears, in recoverable form, the same *information* as was borne by the original signal. We discuss now a number of useful purposes which may be served by frequency translation.

### Frequency Multiplexing

Suppose that we have several different signals, all of which encompass the same spectral range. Let it be required that all these signals be transmitted along a single communications channel in such a manner that, at the receiving end, the signals be separately recoverable and distinguishable from each other. The single channel may be a single pair of wires or the free space that separates one radio antenna from another. Such multiple transmissions, i.e., multiplexing, may be achieved by translating each one of the original signals to a different frequency range. Suppose, say, that one signal is translated to the frequency range  $f'_1$  to  $f'_2$ , the second to the range  $f''_1$  to  $f''_2$ , and so on. If these new frequency ranges do not overlap, then the signal may be separated at the receiving end by appropriate bandpass filters, and the outputs of the filters processed to recover the original signals.

### Practicability of Antennas

When free space is the communications channel, antennas radiate and receive the signal. It turns out that antennas operate effectively only when their dimensions are of the order of magnitude of the wavelength of the signal being transmitted. A signal of frequency 1 kHz (an audio tone) corresponds to a wavelength of 300,000 m, an entirely impractical length. The required length may be reduced to the point of practicability by translating the audio tone to a higher frequency.

### Narrowbanding

Returning to the matter of the antenna, just discussed, suppose that we wanted to transmit an audio signal directly from the antenna, and that the inordinate length of the antenna were no problem. We would still be left with a problem of another type. Let us assume that the audio range extends from, say, 50 to  $10^4$  Hz. The ratio of the highest audio frequency to the lowest is 200. Therefore, an antenna suitable for use at one end of the range would be entirely too short or too long for the other end. Suppose, however, that the audio spectrum were translated so that it occupied the range, say, from  $(10^6 + 50)$  to  $(10^6 + 10^4)$  Hz. Then the ratio of highest to lowest frequency would be only 1.01. Thus the processes of frequency translation may be used to change a "wideband" signal into a "narrowband" signal which may well be more conveniently processed. The terms "wideband" and "narrowband" are being used here to refer not to an absolute range of frequencies but rather to the fractional change in frequency from one band edge to the other.

### Common Processing

It may happen that we may have to process, in turn, a number of signals similar in general character but occupying different spectral ranges. It will then be necessary, as we go from signal to signal, to adjust the frequency range of our processing apparatus to correspond to the frequency range of the signal to be processed. If the processing apparatus is rather elaborate, it may well be wiser to leave the processing apparatus to operate in some fixed frequency range and instead to translate the frequency range of each signal in turn to correspond to this fixed frequency.

## 3.2 A METHOD OF FREQUENCY TRANSLATION

A signal may be translated to a new spectral range by *multiplying* the signal with an auxiliary sinusoidal signal. To illustrate the process, let us consider initially that the signal is sinusoidal in waveform and given by

$$v_m(t) = A_m \cos \omega_m t = A_m \cos 2\pi f_m t \quad (3.2-1a)$$

$$= \frac{A_m}{2} (e^{j\omega_m t} + e^{-j\omega_m t}) = \frac{A_m}{2} (e^{j2\pi f_m t} + e^{-j2\pi f_m t}) \quad (3.2-1b)$$

in which  $A_m$  is the constant amplitude and  $f_m = \omega_m/2\pi$  is the frequency. The two-sided spectral amplitude pattern of this signal is shown in Fig. 3.2-1a. The pattern consists of two lines, each of amplitude  $A_m/2$ , located at  $f = f_m$  and at  $f = -f_m$ . Consider next the result of the multiplication of  $v_m(t)$  with an auxiliary sinusoidal signal

$$v_c(t) = A_c \cos \omega_c t = A_c \cos 2\pi f_c t \quad (3.2-2a)$$

$$= \frac{A_c}{2} (e^{j\omega_c t} + e^{-j\omega_c t}) = \frac{A_c}{2} (e^{j2\pi f_c t} + e^{-j2\pi f_c t}) \quad (3.2-2b)$$

in which  $A_c$  is the constant amplitude and  $f_c$  is the frequency. Using the trigonometric identity  $\cos \alpha \cos \beta = \frac{1}{2} \cos (\alpha + \beta) + \frac{1}{2} \cos (\alpha - \beta)$ , we have for the product  $v_m(t)v_c(t)$

$$v_m(t)v_c(t) = \frac{A_m A_c}{2} [\cos (\omega_c + \omega_m)t + \cos (\omega_c - \omega_m)t] \quad (3.2-3a)$$

$$= \frac{A_m A_c}{4} (e^{j(\omega_c + \omega_m)t} + e^{-j(\omega_c + \omega_m)t} + e^{j(\omega_c - \omega_m)t} + e^{-j(\omega_c - \omega_m)t}) \quad (3.2-3b)$$

The new spectral amplitude pattern is shown in Fig. 3.2-1b. Observe that the two original spectral lines have been *translated*, both in the positive-frequency direction by amount  $f_c$  and also in the negative-frequency direction by the same amount. There are now four spectral components resulting in two sinusoidal waveforms, one of frequency  $f_c + f_m$  and the other of frequency  $f_c - f_m$ . Note that

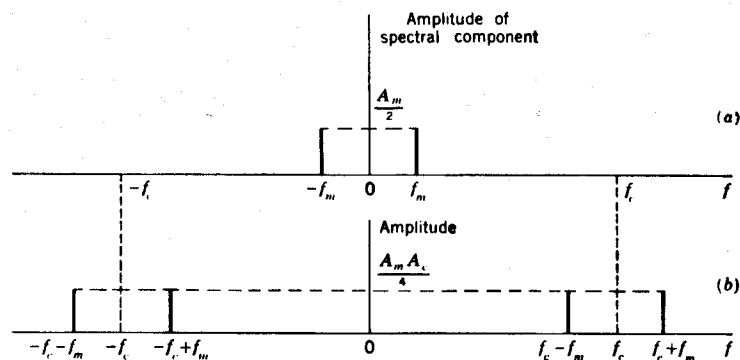


Figure 3.2-1 (a) Spectral pattern of the waveform  $A_m \cos \omega_m t$ . (b) Spectral pattern of the product waveform  $A_m A_c \cos \omega_m t \cos \omega_c t$ .

while the product signal has four spectral components each of amplitude  $A_m A_c/4$ , there are only two frequencies, and the amplitude of each sinusoidal component is  $A_m A_c/2$ .

A generalization of Fig. 3.2-1 is shown in Fig. 3.2-2. Here a signal is chosen which consists of a superposition of four sinusoidal signals, the highest in frequency having the frequency  $f_M$ . Before translation by multiplication, the two-sided spectral pattern displays eight components centered around zero frequency. After multiplication, we find this spectral pattern translated both in the positive and the negative-frequency directions. The 16 spectral components in this two-sided spectral pattern give rise to eight sinusoidal waveforms. While the original signal extends in range up to a frequency  $f_M$ , the signal which results from multiplication has sinusoidal components covering a range  $2f_M$ , from  $f_c - f_M$  to  $f_c + f_M$ .

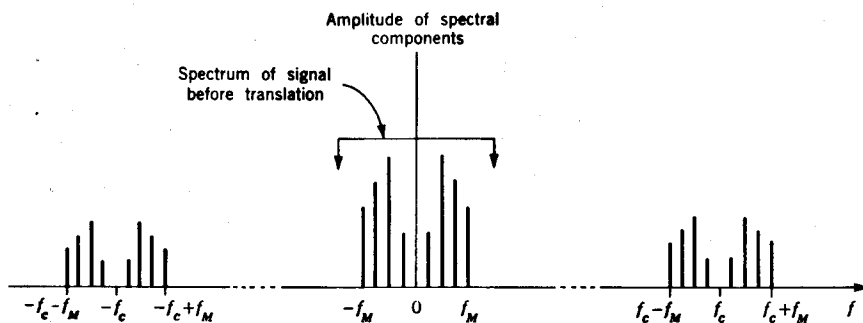


Figure 3.2-2. An original signal consisting of four sinusoids of differing frequencies is translated through multiplication and becomes a signal containing eight frequencies symmetrically arranged about  $f_c$ .

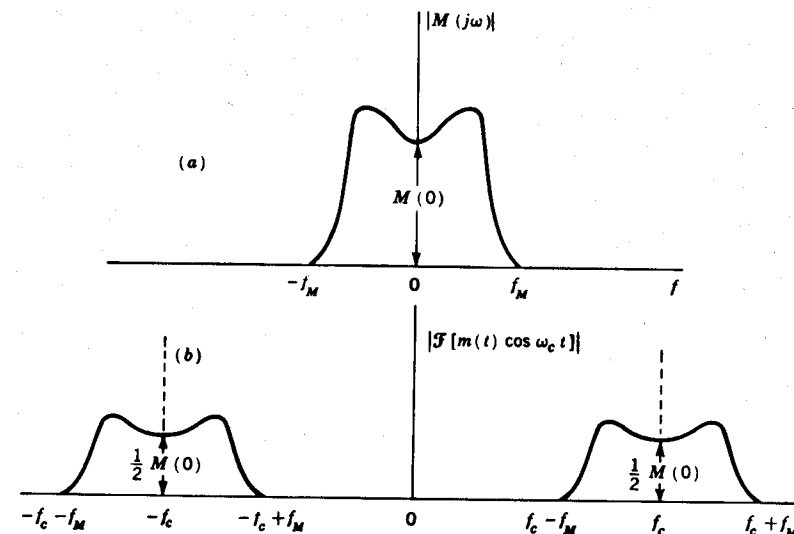


Figure 3.2-3 (a) The spectral density  $|M(j\omega)|$  of a nonperiodic signal  $m(t)$ . (b) The spectral density of  $m(t) \cos 2\pi f_c t$ .

Finally, we consider in Fig. 3.2-3 the situation in which the signal to be translated may not be represented as a superposition of a number of sinusoidal components at sharply defined frequencies. Such would be the case if the signal were of finite energy and nonperiodic. In this case the signal is represented in the frequency domain in terms of its Fourier transform, that is, in terms of its spectral density. Thus let the signal  $m(t)$  be bandlimited to the frequency range 0 to  $f_M$ . Its Fourier transform is  $M(j\omega) = \mathcal{F}[m(t)]$ . The magnitude  $|M(j\omega)|$  is shown in Fig. 3.2-3a. The transform  $M(j\omega)$  is symmetrical about  $f = 0$  since we assume that  $m(t)$  is a real signal. The spectral density of the signal which results when  $m(t)$  is multiplied by  $\cos \omega_c t$  is shown in Fig. 3.2-3b. This spectral pattern is deduced as an extension of the results shown in Figs. 3.2-1 and 3.2-2. Alternatively, we may easily verify (Prob. 3.2-2) that if  $M(j\omega) = \mathcal{F}[m(t)]$ , then

$$\mathcal{F}[m(t) \cos \omega_c t] = \frac{1}{2}[M(j\omega + j\omega_c) + M(j\omega - j\omega_c)] \quad (3.2-4)$$

The spectral range occupied by the original signal is called the *baseband frequency range* or simply the *baseband*. On this basis, the original signal itself is referred to as the *baseband signal*. The operation of multiplying a signal with an auxiliary sinusoidal signal is called *mixing* or *heterodyning*. In the translated signal, the part of the signal which consists of spectral components *above* the auxiliary signal, in the range  $f_c$  to  $f_c + f_M$ , is called the *upper-sideband signal*. The part of the signal which consists of spectral components *below* the auxiliary signal, in the range  $f_c - f_M$  to  $f_c$ , is called the *lower-sideband signal*. The two sideband signals are also referred to as the *sum* and the *difference frequencies*, respec-

tively. The auxiliary signal of frequency  $f_c$  is variously referred to as the *local oscillator signal*, the *mixing signal*, the *heterodyning signal*, or as the *carrier signal*, depending on the application. The student will note, as the discussion proceeds, the various contexts in which the different terms are appropriate.

We may note that the process of translation by multiplication actually gives us something somewhat different from what was intended. Given a signal occupying a baseband, say, from zero to  $f_M$ , and an auxiliary signal  $f_c$ , it would often be entirely adequate to achieve a simple translation, giving us a signal occupying the range  $f_c$  to  $f_c + f_M$ , that is, the upper sideband. We note, however, that translation by multiplication results in a signal that occupies the range  $f_c - f_M$  to  $f_c + f_M$ . This feature of the process of translation by multiplication may, depending on the application, be a nuisance, a matter of indifference, or even an advantage. Hence, this feature of the process is, of itself, neither an advantage nor a disadvantage. It is, however, to be noted that there is no other operation so simple which will accomplish translation.

### 3.3 RECOVERY OF THE BASEBAND SIGNAL

Suppose a signal  $m(t)$  has been translated out of its baseband through multiplication with  $\cos \omega_c t$ . How is the signal to be recovered? The recovery may be achieved by a reverse translation, which is accomplished simply by multiplying the translated signal with  $\cos \omega_c t$ . That such is the case may be seen by drawing spectral plots as in Fig. 3.2-2 or 3.2-3 and noting that the difference-frequency signal obtained by multiplying  $m(t) \cos \omega_c t$  by  $\cos \omega_c t$  is a signal whose spectral range is back at baseband. Alternatively, we may simply note that

$$[m(t) \cos \omega_c t] \cos \omega_c t = m(t) \cos^2 \omega_c t = m(t) \left( \frac{1}{2} + \frac{1}{2} \cos 2\omega_c t \right) \quad (3.3-1a)$$

$$= \frac{m(t)}{2} + \frac{m(t)}{2} \cos 2\omega_c t \quad (3.3-1b)$$

Thus, the baseband signal  $m(t)$  reappears. We note, of course, that in addition to the recovered baseband signal there is a signal whose spectral range extends from  $f_c - f_M$  to  $2f_c + f_M$ . As a matter of practice, this latter signal need cause no difficulty. For most commonly  $f_c \gg f_M$ , and consequently the spectral range of this double-frequency signal and the baseband signal are widely separated. Therefore the double-frequency signal is easily removed by a low-pass filter.

This method of signal recovery, for all its simplicity, is beset by an important inconvenience when applied in a physical communication system. Suppose that the auxiliary signal used for recovery differs in phase from the auxiliary signal used in the initial translation. If this phase angle is  $\theta$ , then, as may be verified Prob. 3.3-1, the recovered baseband waveform will be proportional to  $m(t) \cos \theta$ . Therefore, unless it is possible to maintain  $\theta = 0$ , the signal strength at recovery will suffer. If it should happen that  $\theta = \pi/2$ , the signal will be lost entirely. Or consider, for example, that  $\theta$  drifts back and forth with time. Then in

this case the signal strength will wax and wane, in addition, possibly, to disappearing entirely from time to time.

Alternatively, suppose that the recovery auxiliary signal is not precisely at frequency  $f_c$  but is instead at  $f_c + \Delta f$ . In this case we may verify (Prob. 3.3-2) that the recovered baseband signal will be proportional to  $m(t) \cos 2\pi \Delta f t$ , resulting in a signal which will wax and wane or even be entirely unacceptable if  $\Delta f$  is comparable to, or larger than, the frequencies present in the baseband signal. This latter contingency is a distinct possibility in many an instance, since usually  $f_c \gg f_M$  so that a small percentage change in  $f_c$  will cause a  $\Delta f$  which may be comparable or larger than  $f_M$ . In telephone or radio systems, an offset  $\Delta f \leq 30$  Hz is deemed acceptable.

We note, therefore, that signal recovery using a second multiplication requires that there be available at the recovery point a signal which is precisely *synchronous* with the corresponding auxiliary signal at the point of the first multiplication. In such a *synchronous* or *coherent* system a *fixed* initial phase discrepancy is of no consequence since a simple phase shifter will correct the matter. Similarly it is not essential that the recovery auxiliary signal be sinusoidal (see Prob. 3.3-3). What is essential is that, in any time interval, the number of cycles executed by the two auxiliary-signal sources be the same. Of course, in a physical system, where some signal distortion is tolerable, some lack of synchronism may be allowed.

When the use of a common auxiliary signal is not feasible, it is necessary to resort to rather complicated means to provide a synchronous auxiliary signal at the location of the receiver. One commonly employed scheme is indicated in Fig. 3.3-1. To illustrate the operation of the synchronizer, we assume that the baseband signal is a sinusoidal  $\cos \omega_m t$ . The received signal is  $s_r(t) = A \cos \omega_m t \cos \omega_c t$ , with  $A$  a constant amplitude. This signal  $s_r(t)$  does not have a spectral component at the angular frequency  $\omega_c$ . The output of the squaring circuit is

$$s_r^2(t) = A^2 \cos^2 \omega_m t \cos^2 \omega_c t \quad (3.3-2a)$$

$$= A^2 \left( \frac{1}{2} + \frac{1}{2} \cos 2\omega_m t \right) \left( \frac{1}{2} + \frac{1}{2} \cos 2\omega_c t \right) \quad (3.3-2b)$$

$$= \frac{A^2}{4} \left[ 1 + \frac{1}{2} \cos 2(\omega_c + \omega_m)t + \frac{1}{2} \cos 2(\omega_c - \omega_m)t + \cos 2\omega_m t + \cos 2\omega_c t \right] \quad (3.3-2c)$$

The filter selects the spectral component  $(A^2/4) \cos 2\omega_c t$ , which is then applied to a circuit which divides the frequency by a factor of 2. (See Prob. 3.3-4.) This fre-

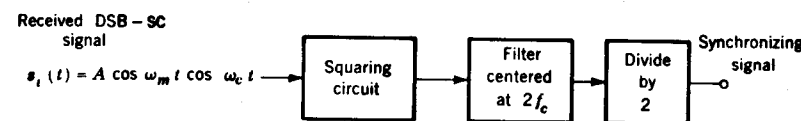


Figure 3.3-1 A simple squaring synchronizer.



quency division may be accomplished by using, for example, a bistable multivibrator. The output of the divider is used to demodulate (multiply) the incoming signal and thereby recover the baseband signal  $\cos \omega_m t$ .

We turn our attention now to a modification of the method of frequency translation, which has the great merit of allowing recovery of the baseband signal by an extremely simple means. This technique is called *amplitude modulation*.

### 3.4 AMPLITUDE MODULATION

A frequency-translated signal from which the baseband signal is easily recoverable is generated by adding, to the product of baseband and carrier, the carrier signal itself. Such a signal is shown in Fig. 3.4-1. Figure 3.4-1a shows the carrier signal with amplitude  $A_c$ , in Fig. 3.4-1b we see the baseband signal. The translated signal (Fig. 3.4-1c) is given by

$$v(t) = A_c[1 + m(t)] \cos \omega_c t \quad (3.4-1)$$

We observe, from Eq. (3.4-1) as well as from Fig. 3.4-1c, that the resultant waveform is one in which the carrier  $A_c \cos \omega_c t$  is *modulated in amplitude*. The process of generating such a waveform is called *amplitude modulation*, and a communication system which employs such a method of frequency translation is called an *amplitude-modulation system*, or *AM* for short. The designation "carrier" for the auxiliary signal  $A_c \cos \omega_c t$  seems especially appropriate in the present connection since this signal now "carries" the baseband signal as its envelope. The term "carrier" probably originated, however, in the early days of radio when this relatively high-frequency signal was viewed as the *messenger* which actually "carried" the baseband signal from one antenna to another.

The very great merit of the amplitude-modulated carrier signal is the ease with which the baseband signal can be recovered. The recovery of the baseband signal, a process which is referred to as *demodulation* or *detection*, is accomplished with the simple circuit of Fig. 3.4-2a, which consists of a diode  $D$  and the resistor-capacitor  $RC$  combination. We now discuss the operation of this circuit briefly and qualitatively. For simplicity, we assume that the amplitude-modulated carrier which is applied at the input terminals is supplied by a voltage source of zero internal impedance. We assume further that the diode is ideal, i.e., of zero or infinite resistance, depending on whether the diode current is positive or the diode voltage negative.

Let us initially assume that the input is of fixed amplitude and that the resistor  $R$  is not present. In this case, the capacitor charges to the peak positive voltage of the carrier. The capacitor holds this peak voltage, and the diode would not again conduct. Suppose now that the input-carrier amplitude is increased. The diode again conducts, and the capacitor charges to the new higher carrier peak. In order to allow the capacitor voltage to follow the carrier peaks when the carrier amplitude is decreasing, it is necessary to include the resistor  $R$ , so that the capacitor may discharge. In this case the capacitor voltage  $v_c$  has the form

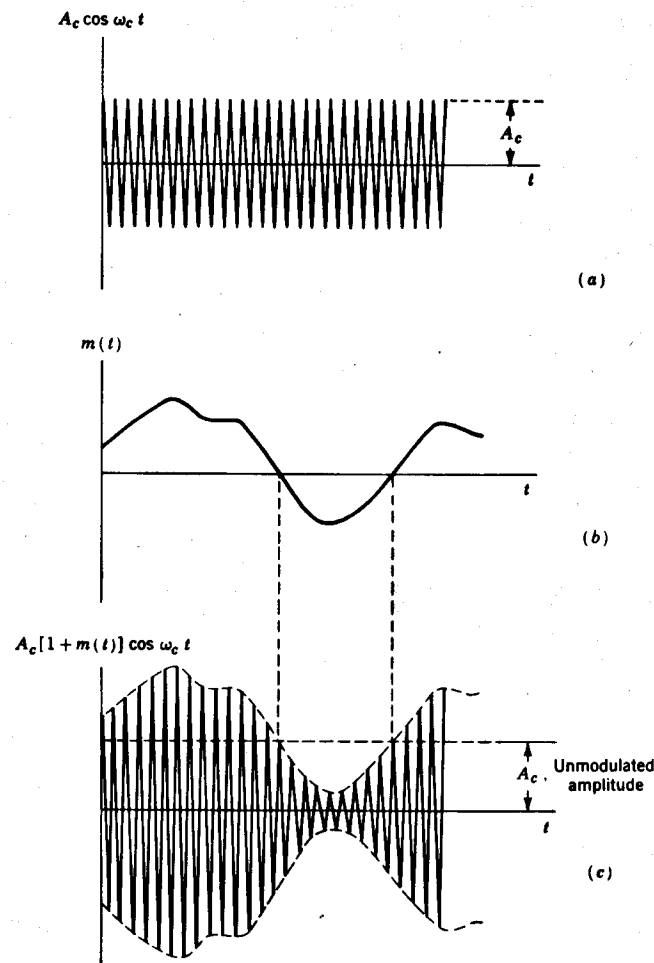


Figure 3.4-1 (a) A sinusoidal carrier. (b) A modulating waveform. (c) The sinusoidal carrier in (a) modulated by the waveform in (b).

shown in Fig. 3.4-2b. The capacitor charges to the peak of each carrier cycle and decays slightly between cycles. The time constant  $RC$  is selected so that the change in  $v_c$  between cycles is at least equal to the decrease in carrier amplitude between cycles. This constraint on the time constant  $RC$  is explored in Probs. 3.4-1 and 3.4-2.

It is seen that the voltage  $v_c$  follows the carrier envelope except that  $v_c$  also has superimposed on it a sawtooth waveform of the carrier frequency. In Fig. 3.4-2b the discrepancy between  $v_c$  and the envelope is greatly exaggerated. In

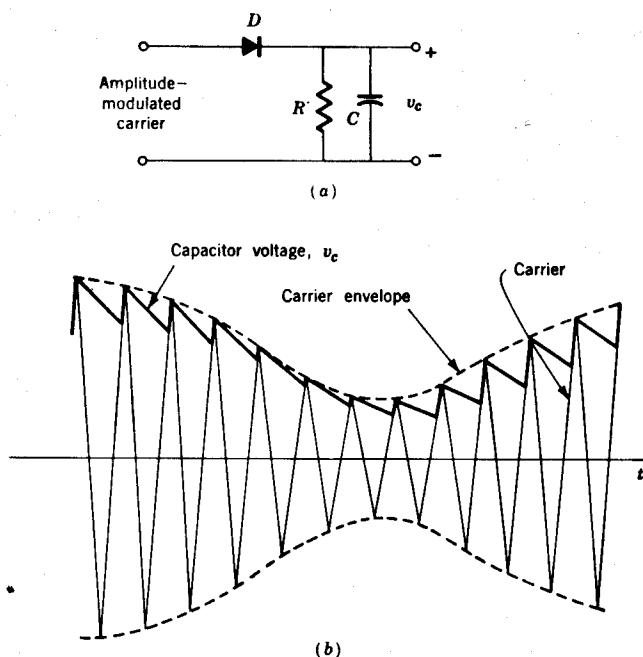


Figure 3.4-2 (a) A demodulator for an AM signal. (b) Input waveform and output voltage  $v_c$  across capacitor.

practice, the normal situation is one in which the time interval between carrier cycles is extremely small in comparison with the time required for the envelope to make a sizeable change. Hence  $v_c$  follows the envelope much more closely than is suggested in the figure. Further, again because the carrier frequency is ordinarily much higher than the highest frequency of the modulating signal, the sawtooth distortion of the envelope waveform is very easily removed by a filter.

### 3.5 MAXIMUM ALLOWABLE MODULATION

If we are to avail ourselves of the convenience of demodulation by the use of the simple diode circuit of Fig. 3.4-2a, we must limit the extent of the modulation of the carrier. That such is the case may be seen from Fig. 3.5-1. In Fig. 3.5-1a is shown a carrier modulated by a sinusoidal signal. It is apparent that the envelope of the carrier has the waveshape of the modulating signal. The modulating signal is sinusoidal; hence  $m(t) = m \cos \omega_m t$ , where  $m$  is a constant. Equation (3.4-1) becomes

$$v(t) = A_c(1 + m \cos \omega_m t) \cos \omega_c t \quad (3.5-1)$$

In Fig. 3.5-1b we have shown the situation which results when, in Eq. (3.5-1), we adjust  $m > 1$ . Observe now that the diode demodulator which yields as an output the positive envelope (a negative envelope if the diode is reversed) will not reproduce the sinusoidal modulating waveform. In this latter case, where  $m > 1$ , we may recover the modulating waveform but not with the diode demodulator. Recovery would require the use of a coherent demodulation scheme such as was employed in connection with the signal furnished by a multiplier.

It is therefore necessary to restrict the excursion of the modulating signal in the direction of decreasing carrier amplitude to the point where the carrier amplitude is just reduced to zero. No such similar restriction applies when the modulation is increasing the carrier amplitude. With sinusoidal modulation, as in Eq. (3.5-1), we require that  $|m| \leq 1$ . More generally in Eq. (3.4-1) we require that the maximum negative excursion of  $m(t)$  be  $-1$ .

The extent to which a carrier has been amplitude-modulated is expressed in terms of a *percentage modulation*. Let  $A_c$ ,  $A_c(\max)$ , and  $A_c(\min)$ , respectively, be

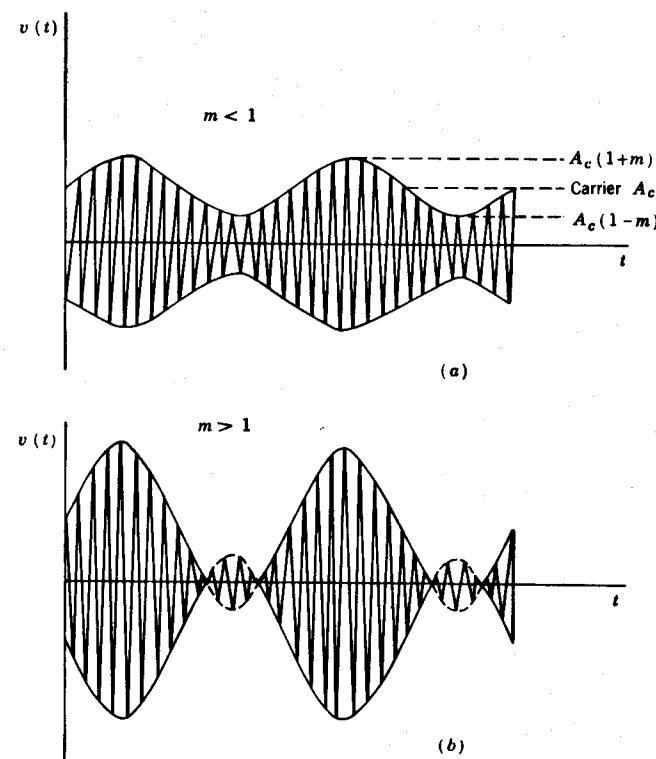


Figure 3.5-1 (a) A sinusoidally modulated carrier ( $m < 1$ ). (b) A carrier overmodulated ( $m > 1$ ) by a sinusoidal modulating waveform.

the unmodulated carrier amplitude and the maximum and minimum carrier levels. Then if the modulation is symmetrical, the percentage modulation is defined as  $P$ , given by

$$\frac{P}{100\%} = \frac{A_c(\max) - A_c}{A_c} = \frac{A_c - A_c(\min)}{A_c} = \frac{A_c(\max) - A_c(\min)}{2A_c} \quad (3.5-2)$$

In the case of sinusoidal modulation, given by Eq. (3.5-1) and shown in Fig. 3.5-1a,  $P = m \times 100$  percent.

Having observed that the signal  $m(t)$  may be recovered from the waveform  $A_c[1 + m(t)] \cos \omega_c t$  by the simple circuit of Fig. 3.4-2a, it is of interest to note that a similar easy recovery of  $m(t)$  is not possible from the waveform  $m(t) \cos \omega_c t$ . That such is the case is to be seen from Fig. 3.5-2. Figure 3.5-2a shows the carrier signal. The modulation or baseband signal  $m(t)$  is shown in Fig.

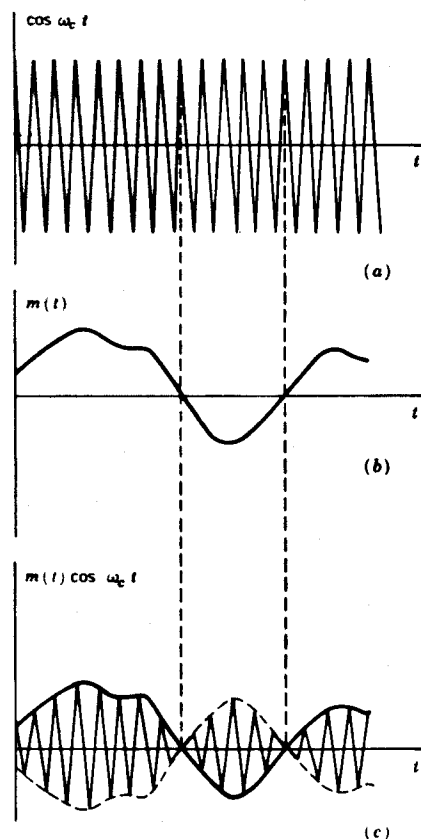


Figure 3.5-2 (a) A carrier  $\cos \omega_c t$ . (b) A baseband signal  $m(t)$ . (c) The product  $m(t) \cos \omega_c t$  and its envelope.

3.5-2b, and the product  $m(t) \cos \omega_c t$  is shown in Fig. 3.5-2c. We note that the envelope in Fig. 3.5-2c has the waveform not of  $m(t)$  but rather of  $|m(t)|$ , the absolute value of  $m(t)$ . Observe the reversal of phase of the carrier in Fig. 3.5-2c whenever  $m(t)$  passes through zero.

### 3.6 THE SQUARE-LAW DEMODULATOR

An alternative method of recovering the baseband signal which has been superimposed as an amplitude modulation on a carrier is to pass the AM signal through a nonlinear device. Such demodulation is illustrated in Fig. 3.6-1. We assume here for simplicity that the device has a square-law relationship between input signal  $x$  (current or voltage) and output signal  $y$  (current or voltage). Thus  $y = kx^2$ , with  $k$  a constant. Because of the nonlinearity of the transfer characteristic of the device, the output response is different for positive and for negative excursions of the carrier away from the quiescent operating point  $O$  of the device.

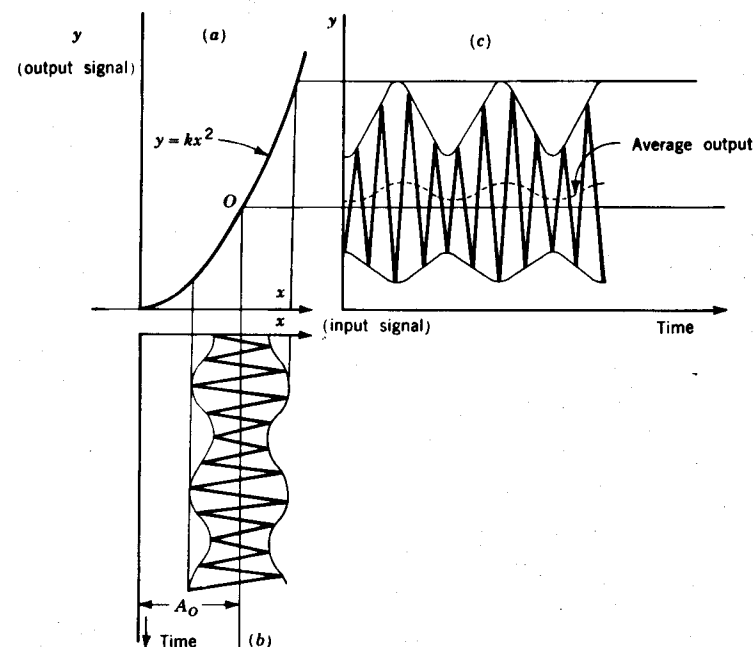


Figure 3.6-1 Illustrating the operation of a square-law demodulator. The output is the value of  $y$  averaged over many carrier cycles.

As a result, and as is shown in Fig. 3.6-1c, the output, when averaged over a time which encompasses many carrier cycles but only a very small part of the modulation cycle, has the waveshape of the envelope.

The applied signal is

$$x = A_o + A_c[1 + m(t)] \cos \omega_c t \quad (3.6-1)$$

Thus the output of the squaring circuit is

$$y = k\{A_o + A_c[1 + m(t)] \cos \omega_c t\}^2 \quad (3.6-2)$$

Squaring, and dropping dc terms as well as terms whose spectral components are located near  $\omega_c$  and  $2\omega_c$ , we find that the output signal  $s_o(t)$ , that is, the signal output of a low-pass filter located after the squaring circuit, is

$$s_o(t) = kA_c^2[m(t) + \frac{1}{2}m^2(t)] \quad (3.6-3)$$

Observe that the modulation  $m(t)$  is indeed recovered but that  $m^2(t)$  appears as well. Thus the total recovered signal is a distorted version of the original modulation. The distortion is small, however, if  $\frac{1}{2}m^2(t) \ll |m(t)|$  or if  $|m(t)| \ll 2$ .

There are two points of interest to be noted in connection with the type of demodulation described here; the first is that the demodulation does not depend on the nonlinearity being square-law. Any type of nonlinearity which does not have odd-function symmetry with respect to the initial operating point will similarly accomplish demodulation. The second point is that even when demodulation is not intended, such demodulation may appear incidentally when the modulated signal is passed through a system, say, an amplifier, which exhibits some nonlinearity.

### 3.7 SPECTRUM OF AN AMPLITUDE-MODULATED SIGNAL

The spectrum of an amplitude-modulated signal is similar to the spectrum of a signal which results from multiplication except, of course, that in the former case a carrier of frequency  $f_c$  is present. If in Eq. (3.4-1)  $m(t)$  is the superposition of three sinusoidal components  $m(t) = m_1 \cos \omega_1 t + m_2 \cos \omega_2 t + m_3 \cos \omega_3 t$ , then the (one-sided) spectrum of this baseband signal appears as at the left in Fig. 3.7-1a. The spectrum of the modulated carrier is shown at the right. The spectral lines at the sum frequencies  $f_c + f_1$ ,  $f_c + f_2$ , and  $f_c + f_3$  constitute the *upper-sideband* frequencies. The spectral lines at the difference frequencies constitute the *lower sideband*.

The spectrum of the baseband signal and modulated carrier are shown in Fig. 3.7-1b for the case of a bandlimited nonperiodic signal of finite energy. In this figure the ordinate is the spectral density, i.e., the magnitude of the Fourier transform rather than the spectral amplitude, and consequently the carrier is represented by an impulse.

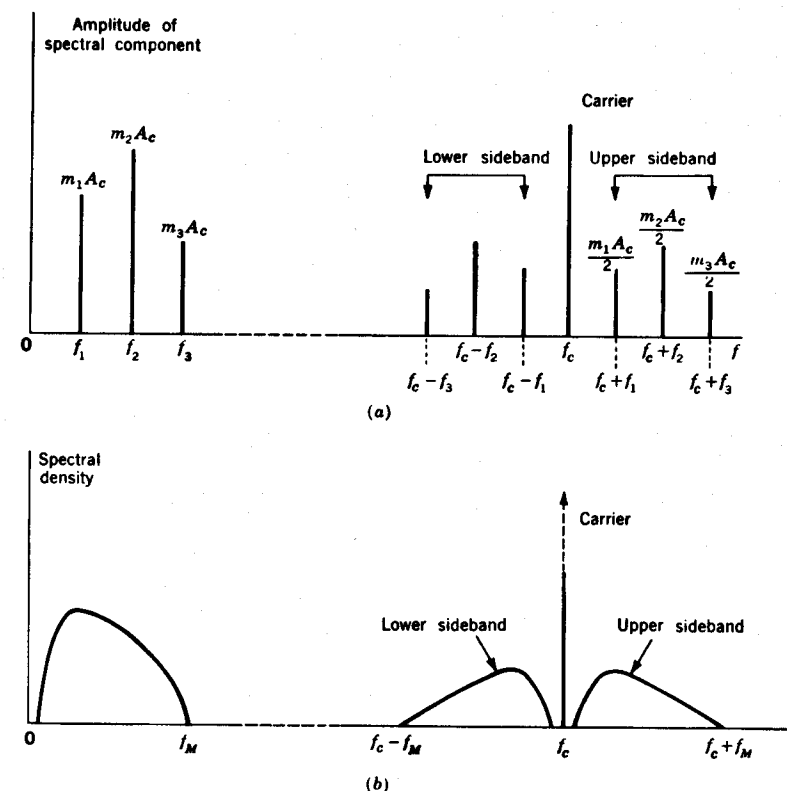


Figure 3.7-1 (a) At left the one-sided spectrum of  $m(t)A_c$ , where  $m(t)$  has three spectral components. At right the spectrum of  $A_c[1 + m(t)] \cos 2\pi f_c t$ . (b) Same as in (a) except  $m(t)$  is a nonperiodic signal and the vertical axis is spectral density rather than spectral amplitude.

### 3.8 MODULATORS AND BALANCED MODULATORS

We have described a "multiplier" as a device that yields as an output a signal which is the product of two input signals. Actually no simple physical device now exists which yields the product alone. On the contrary, all such devices yield, at a minimum, not only the product but the input signals themselves. Suppose, then, that such a device has as inputs a carrier  $\cos \omega_c t$  and a modulating baseband signal  $m(t)$ . The device output will then contain the product  $m(t) \cos \omega_c t$  and also the signals  $m(t)$  and  $\cos \omega_c t$ . Ordinarily, the baseband signal will be bandlimited to a frequency range very much smaller than  $f_c = \omega_c/2\pi$ . Suppose, for example, that the baseband signal extends from zero frequency to 1000 Hz, while  $f_c = 1$  MHz. In this case, the carrier and its sidebands extend from 999,000 to 1,001,000 Hz, and the baseband signal is easily removed by a filter.

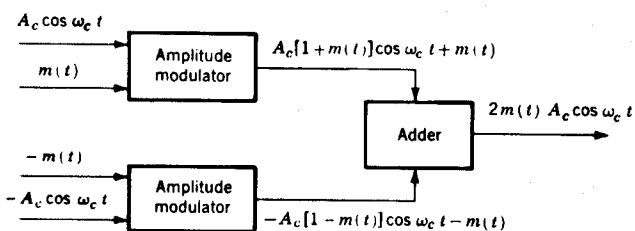


Figure 3.8-1 Showing how the outputs of two amplitude modulators are combined to produce a double-sideband suppressed-carrier output.

The overall result is that the devices available for multiplication yield an output carrier as well as the lower- and upper-sideband signals. The output is therefore an amplitude-modulated signal. If we require the product signal alone, we must take steps to cancel or *suppress* the carrier. Such a suppression may be achieved by adding, to the amplitude-modulated signal, a signal of carrier frequency equal in amplitude but opposite in phase to the carrier of the amplitude-modulated signal. Under these circumstances only the sideband signals will remain. For this reason, a product signal is very commonly referred to as a *double-sideband suppressed-carrier* signal, abbreviated DSB-SC.

An alternative arrangement for carrier suppression is shown in Fig. 3.8-1. Here two *physical* multipliers are used which are labeled in the diagram as *amplitude modulators*. The carrier inputs to the two modulators are of reverse polarity, as are the modulating signals. The modulator outputs are added with consequent suppression of the carrier. We observe a cancellation not only of the carrier but of the baseband signal  $m(t)$  as well. This last feature is not of great import, since, as noted previously, the baseband signal is easily eliminated by a filter. We note that the product terms of the two modulators reinforce. The arrangement of Fig. 3.8-1 is called a *balanced modulator*.

### 3.9 SINGLE-SIDEBAND MODULATION

We have seen that the baseband signal may be recovered from a double-sideband suppressed-carrier signal by multiplying a second time, with the same carrier. It can also be shown that the baseband signal can be recovered in a similar manner even if only one sideband is available. For suppose a spectral component of the baseband signal is multiplied by a carrier  $\cos \omega_c t$ , giving rise to an upper sideband at  $\omega_c + \omega$  and a lower sideband at  $\omega_c - \omega$ . Now let us assume that we have filtered out one sideband and are left with, say, only the upper sideband at  $\omega_c + \omega$ . If now this sideband signal is again multiplied by  $\cos \omega_c t$ , we shall generate a signal at  $2\omega_c + \omega$  and the original baseband spectral component at  $\omega$ . If we had used the lower sideband at  $\omega_c - \omega$ , the second multiplication would have yielded a signal at  $2\omega_c - \omega$  and again restored the baseband spectral component.

Since it is possible to recover the baseband signal from a single sideband, there is an obvious advantage in doing so, since spectral space is used more economically. In principle, two single-sideband (abbreviated SSB) communications systems can now occupy the spectral range previously occupied by a single amplitude-modulation system or a double-sideband suppressed-carrier system.

The baseband signal may *not* be recovered from a single-sideband signal by the use of a diode modulator. That such is the case is easily seen by considering, for example, that the modulating signal is a sinusoid of frequency  $f$ . In this case the single-sideband signal will consist also of a single sinusoid of frequency, say,  $f_c + f$ , and there is no amplitude variation at all at the baseband frequency to which the diode modulator can respond.

Baseband recovery is achieved at the receiving end of the single-sideband communications channel by heterodyning the received signal with a local carrier signal which is synchronous (coherent) with the carrier used at the transmitting end to generate the sideband. As in the double-sideband case it is necessary, in principle, that the synchronism be exact and, in practice, that synchronism be maintained to a high order of precision. The effect of a lack of synchronism is different in a double-sideband system and in a single-sideband system. Suppose that the carrier received is  $\cos \omega_c t$  and that the local carrier is  $\cos(\omega_c t + \theta)$ . Then with DSB-SC, as noted in Sec. 3.3, the spectral component  $\cos \omega t$  will, upon demodulation, reappear as  $\cos \omega t \cos \theta$ . In SSB, on the other hand, the spectral component,  $\cos \omega t$  will reappear (Prob. 3.9-2) in the form  $\cos(\omega t - \theta)$ . Thus, in one case a phase offset in carriers affects the amplitude of the recovered signal and, for  $\phi = \pi/2$ , may result in a total loss of the signal. In the other case the offset produces a phase change but not an amplitude change.

Alternatively, let the local oscillator carrier have an angular frequency offset  $\Delta\omega$  and so be of the form  $\cos(\omega_c + \Delta\omega)t$ . Then as already noted, in DSB-SC, the recovered signal has the form  $\cos \omega t \cos \Delta\omega t$ . In SSB, however, the recovered signal will have the form  $\cos(\omega + \Delta\omega)t$ . Thus, in one case the recovered spectral component  $\cos \omega t$  reappears with a "warble," that is, an amplitude fluctuation at the rate  $\Delta\omega$ . In the other case the amplitude remains fixed, but the frequency of the recovered signal is in error by amount  $\Delta\omega$ .

A phase offset between the received carrier and the local oscillator will cause distortion in the recovered baseband signal. In such a case each spectral component in the baseband signal will, upon recovery, have undergone the *same* phase shift. Fortunately, when SSB is used to transmit voice or music, such *phase distortion* does not appear to be of major consequence, because the human ear seems to be insensitive to the phase distortion.

A frequency offset between carriers in amount of  $\Delta f$  will cause each recovered spectral component of the baseband signal to be in error by the same amount  $\Delta f$ . Now, if it had turned out that the frequency error were proportional to the frequency of the spectral component itself, then the recovered signal would sound like the original signal except that it would be at a higher or lower pitch. Such, however, is not the case, since the frequency error is fixed. Thus frequencies in the original signal which were harmonically related will no longer be so related after

recovery. The overall result is that a frequency offset between carriers adversely affects the intelligibility of spoken communication and is not well tolerated in connection with music. As a matter of experience, it turns out that an error  $\Delta f$  of less than 30 Hz is acceptable to the ear.

The need to keep the frequency offset  $\Delta f$  between carriers small normally imposes severe restrictions on the frequency stabilities of the carrier signal generators at both ends of the communications system. For suppose that we require to keep  $\Delta f$  to 10 Hz or less, and that our system uses a carrier frequency of 10 MHz. Then the sum of the frequency drift in the two carrier generators may not exceed 1 part in  $10^6$ . The required equality in carrier frequency may be maintained through the use of quartz crystal oscillators using crystals cut for the same frequency at transmitter and receiver. The receiver must use as many crystals (or equivalent signals derived from crystals) as there are channels in the communications system.

It is also possible to tune an SSB receiver manually and thereby reduce the frequency offset. To do this, the operator manually adjusts the frequency of the receiver carrier generator until the received signal sounds "normal." Experienced operators are able to tune correctly to within 10 or 20 Hz. However, because of oscillator drift, such tuning must be readjusted periodically.

When the carrier frequency is very high, even quartz crystal oscillators may be hard pressed to maintain adequate stability. In such cases it is necessary to transmit the carrier itself along with the sideband signal. At the receiver the carrier may be separated by filtering and used to synchronize a local carrier generator. When used for such synchronization, the carrier is referred to as a "pilot carrier" and may be transmitted at a substantially reduced power level.

It is interesting to note that the squaring circuit used to recover the frequency and phase information of the DSB-SC system cannot be used here. In any event, it is clear that a principal complication in the way of more widespread use of single sideband is the need for supplying an accurate carrier frequency at the receiver.

### 3.10 METHODS OF GENERATING AN SSB SIGNAL

#### Filter Method

A straightforward method of generating an SSB signal is illustrated in Fig. 3.10-1. Here the baseband signal and a carrier are applied to a balanced modulator. The output of the balanced modulator bears both the upper- and lower-sideband signals. One or the other of these signals is then selected by a filter. The filter is a bandpass filter whose passband encompasses the frequency range of the sideband selected. The filter must have a cutoff sharp enough to separate the selected sideband from the other sideband. The frequency separation of the sidebands is twice the frequency of the lowest frequency spectral components of the baseband signal. Human speech contains spectral components as low as about 70 Hz.

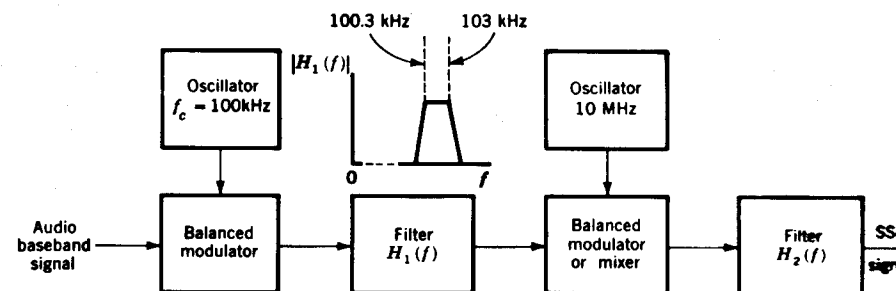


Figure 3.10-1 Block diagram of the filter method of generating a single-sideband signal.

However, to alleviate the sideband filter selectivity requirements in an SSB system, it is common to limit the lower spectral limit of speech to about 300 Hz. It is found that such restriction does not materially affect the intelligibility of speech. Similarly, it is found that no serious distortion results if the upper limit of the speech spectrum is cut off at about 3000 Hz. Such restriction is advantageous for the purpose of conserving bandwidth. Altogether, then, a typical sideband filter has a passband which, measured from  $f_c$ , extends from about 300 to 3000 Hz and in which range its response is quite flat. Outside this passband the response falls off sharply, being down about 40 dB at 4000 Hz and rejecting the unwanted sideband also be at least 40 dB. The filter may also serve, further, to suppress the carrier itself. Of course, in principle, no carrier should appear at the output of a balanced modulator. In practice, however, the modulator may not balance exactly, and the precision of its balance may be subject to some variation with time. Therefore, even if a pilot carrier is to be transmitted, it is well to suppress it at the output of the modulator and to add it to the signal at a later point in a controllable manner.

Now consider that we desire to generate an SSB signal with a carrier of, say, 10 MHz. Then we require a passband filter with a selectivity that provides 40 dB of attenuation within 600 Hz at a frequency of 10 MHz, a percentage frequency change of 0.006 percent. Filters with such sharp selectivity are very elaborate and difficult to construct. For this reason, it is customary to perform the translation of the baseband signal to the final carrier frequency in several stages. Two such stages of translation are shown in Fig. 3.10-1. Here we have selected the first carrier to be of frequency 100 kHz. The upper sideband, say, of the output of the balanced modulator ranges from 100.3 to 103 kHz. The filter following the balanced modulator which selects this upper sideband need now exhibit a selectivity of only a hundredth of the selectivity (40 dB in 0.6 percent frequency change) required in the case of a 10-MHz carrier. Now let the filter output be applied to a second balanced modulator, supplied this time with a 10-MHz carrier. Let us again select the upper sideband. Then the second filter must provide 40 dB of attenuation in a frequency range of 200.6 kHz, which is nominally 2 percent of the carrier frequency.

We have already noted that the simplest physical frequency-translating device is a multiplier or mixer, while a balanced modulator is a balanced arrangement of two mixers. A mixer, however, has the disadvantage that it presents at its output not only sum and difference frequencies but the input frequencies as well. Still, when it is feasible to discriminate against these input signals, there is a merit of simplicity in using a mixer rather than a balanced modulator. In the present case, if the second frequency-translating device in Fig. 3.10-1 were a mixer rather than a multiplier, then in addition to the upper and lower sidebands, the output would contain a component encompassing the range 100.3 to 103 kHz as well as the 10-MHz carrier. The range 100.3 to 103 kHz is well out of the range of the second filter intended to pass the range 10,100,300 to 10,103,000 Hz. And it is realistic to design a filter which will suppress the 10-MHz carrier, since the carrier frequency is separated from the lower edge of the upper sideband (10,100,300) by nominally a 1-percent frequency change.

Altogether, then, we note in summary that when a single-sideband signal is to be generated which has a carrier in the megahertz or tens-of-megahertz range, the frequency translation is to be done in more than one stage—frequently two but not uncommonly three. If the baseband signal has spectral components in the range of hundreds of hertz or lower (as in an audio signal), the first stage invariably employs a balanced modulator, while succeeding stages may use mixers.

### Phasing Method

An alternative scheme for generating a single-sideband signal is shown in Fig. 3.10-2. Here two balanced modulators are employed. The carrier signals of angular frequency  $\omega_c$  which are applied to the modulators differ in phase by  $90^\circ$ .

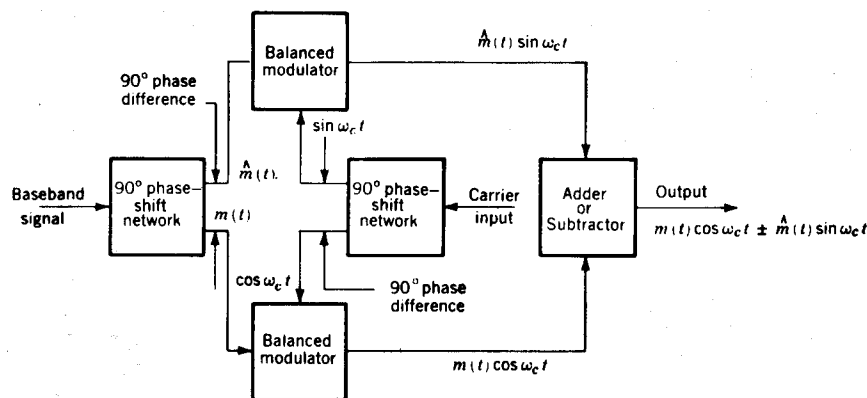


Figure 3.10-2 A method of generating a single-sideband signal using balanced modulators and phase shifters.

Similarly the baseband signal, before application to the modulators, is passed through a  $90^\circ$  phase-shifting network so that there is a  $90^\circ$  phase shift between any spectral component of the baseband signal applied to one modulator and the like-frequency component applied to the other modulator.

To see most simply how the arrangement of Fig. 3.10-2 operates, let us assume that the baseband signal is sinusoidal and appears at the input to one modulator as  $\cos \omega_m t$  and hence as  $\sin \omega_m t$  at the other. Also, let the carrier be  $\cos \omega_c t$  at one modulator and  $\sin \omega_c t$  at the other. Then the outputs of the balanced modulators (multipliers) are

$$\cos \omega_m t \cos \omega_c t = \frac{1}{2} [\cos (\omega_c - \omega_m) t + \cos (\omega_c + \omega_m) t] \quad (3.10-1)$$

$$\sin \omega_m t \sin \omega_c t = \frac{1}{2} [\cos (\omega_c - \omega_m) t - \cos (\omega_c + \omega_m) t] \quad (3.10-2)$$

If these waveforms are added, the lower sideband results; if subtracted, the upper sideband appears at the output. In general, if the modulation  $m(t)$  is given by

$$m(t) = \sum_{i=1}^m A_i \cos (\omega_i t + \theta_i) \quad (3.10-3)$$

then, using Fig. 3.10-2, we see that the output of the SSB modulator is in general

$$m(t) \cos \omega_c t \pm \hat{m}(t) \sin \omega_c t \quad (3.10-4)$$

where

$$\hat{m}(t) \equiv \sum_{i=1}^m A_i \sin (\omega_i t + \theta_i) \quad (3.10-5)$$

The single-sideband generating system of Fig. 3.10-2 generally enjoys less popularity than does the filter method. The reason for this lack of favor is that the present phasing method requires, for satisfactory operation, that a number of constraints be rather precisely met if the carrier and one sideband are adequately to be suppressed. It is required that each modulator be rather carefully balanced to suppress the carrier. It requires also that the baseband signal phase-shifting network provide to the modulators signals in which equal frequency spectral components are of exactly equal amplitude and differ in phase by precisely  $90^\circ$ . Such a network is difficult to construct for a baseband signal which extends over many octaves. It is also required that each modulator display equal sensitivity to the baseband signal. Finally, the carrier phase-shift network must provide exactly  $90^\circ$  of phase shift. If any of these constraints is not satisfied, the suppression of the rejected sideband and of the carrier will suffer. The effect on carrier and sideband suppression due to a failure precisely to meet these constraints is explored in Probs. 3.10-3 and 3.10-4. Of course, in any physical system a certain level of carrier and rejected sideband is tolerable. Still, there seems to be a general inclination to achieve a single sideband by the use of passive filters rather than by a method which requires many exactly maintained balances in passive and active circuits. There is an alternative single-sideband generating scheme<sup>2</sup> which avoids the need for a wideband phase-shifting network but which uses four balanced modulators.

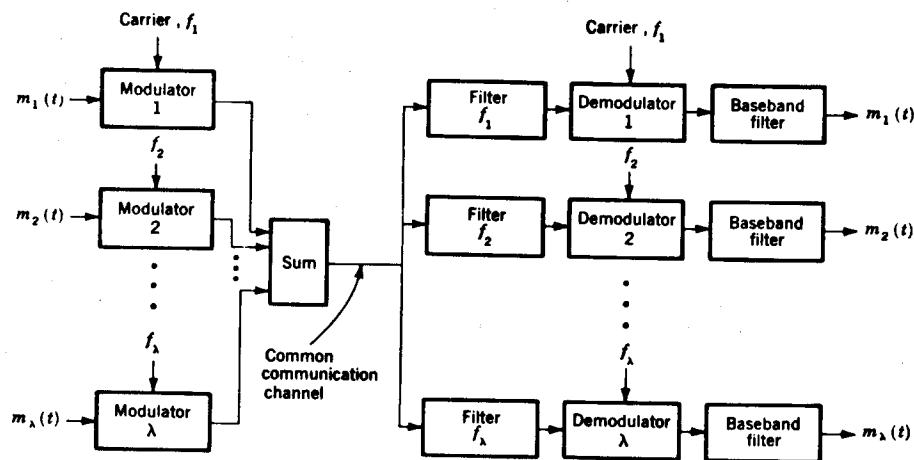


Figure 3.13-1 Multiplexing many baseband signals over a single communications channel.

**multiplexing.** As a matter of fact, to facilitate this separation of the individual signals, the carrier frequencies are selected to leave a comfortable margin (guard band) between the limit of one frequency range and the beginning of the next. The combined output of all the modulators, i.e., the *composite* signal, is applied to a common communications channel. In the case of radio transmission, the channel is free space, and coupling to the channel is made by means of an antenna. In other cases wires are used.

At the receiving end the composite signal is applied to each of a group of bandpass filters whose passbands are in the neighborhood  $f_1, f_2, \dots, f_\lambda$ . The filter  $f_1$  is a bandpass filter which passes only the spectral range of the output of modulator 1 and similarly for the other bandpass filters. The signals have thus been separated. They are then applied to individual demodulators which extract the baseband signals from the carrier. The carrier inputs to the demodulators are required only for synchronous demodulation and are not used otherwise.

The final operation indicated in Fig. 3.13-1 consists in passing the demodulator output through a baseband filter. The baseband filter is a low-pass filter with cutoff at the frequency  $f_M$  to which the baseband signal is limited. This baseband filter will pass, without modification, the baseband signal output of the modulator and in this sense serves no function in the system as described up to the present point. We shall, however, see in Chaps. 8 and 9 that such baseband filters are essential to suppress the noise which invariably accompanies the signal.

## REFERENCES

1. Bell Telephone Laboratories: "Transmission Systems for Communications," Western Electric Company, Tech. Pub., Winston-Salem, N.C., 1964.

2. Norgaard, D. E.: A Third Method of Generation and Detection of Single-sideband Signals, *Proc. IRE*, December, 1956.
3. Voelcker, H.: Demodulation of Single-sideband Signals Via Envelope Detection, *IEEE Trans. on Communication Technology*, pp. 22-30, February, 1966.

## PROBLEMS

3.2-1. A signal  $v_m(t)$  is bandlimited to the frequency range 0 to  $f_M$ . It is frequency-translated by multiplying it by the signal  $v_c(t) = \cos 2\pi f_c t$ . Find  $f_c$  so that the bandwidth of the translated signal is 1 percent of the frequency  $f_c$ .

3.2-2. The Fourier transform of  $m(t)$  is  $\mathcal{F}[m(t)] \equiv M(f)$ . Show that

$$\mathcal{F}[m(t) \cos 2\pi f_c t] = \frac{1}{2}[M(f + f_c) + M(f - f_c)]$$

3.2-3. The signals

$$v_1(t) = 2 \cos \omega_1 t + \cos 2\omega_1 t$$

and

$$v_2(t) = \cos \omega_2 t + 2 \cos 2\omega_2 t$$

are multiplied. Plot the resultant amplitude-frequency characteristic, assuming that  $\omega_2 > 2\omega_1$ , but is not a harmonic of  $\omega_1$ . Repeat for  $\omega_2 = 2\omega_1$ .

3.3-1. The baseband signal  $m(t)$  in the frequency-translated signal  $v(t) = m(t) \cos 2\pi f_c t$  is recovered by multiplying  $v(t)$  by the waveform  $\cos(2\pi f_c t + \theta)$ .

(a) The product waveform is transmitted through a low-pass filter which rejects the double-frequency signal. What is the output signal of the filter?

(b) What is the maximum allowable value for the phase  $\theta$  if the recovered signal is to be 90 percent of the maximum possible value?

(c) If the baseband signal  $m(t)$  is bandlimited to 10 kHz, what is the minimum value of  $f_c$  for which it is possible to recover  $m(t)$  by filtering the product waveform  $v(t) \cos(2\pi f_c t + \theta)$ ?

3.3-2. The baseband signal  $m(t)$  in the frequency-translated signal  $v(t) = m(t) \cos 2\pi f_c t$  is recovered by multiplying  $v(t)$  by the waveform  $\cos 2\pi(f_c + \Delta f)t$ . The product waveform is transmitted through a low-pass filter which rejects the double-frequency signal. Find the output signal of the filter.

3.3-3. (a) The baseband signal  $m(t)$  in the frequency-translated signal  $v(t) = m(t) \cos 2\pi f_c t$  is to be recovered. There is available a waveform  $p(t)$  which is periodic with period  $1/f_c$ . Show that  $m(t)$  may be recovered by appropriately filtering the product waveform  $p(t)v(t)$ .

(b) Show that  $m(t)$  may be recovered as well if the periodic waveform has a period  $n/f_c$ , where  $n$  is an integer. Assume  $m(t)$  bandlimited to the frequency range from 0 to 5 kHz and let  $f_c = 1$  MHz. Find the largest  $n$  which will allow  $m(t)$  to be recovered. Are all periodic waveforms acceptable?

3.3-4. The signal  $m(t)$  in the DSB-SC signal  $v(t) = m(t) \cos(\omega_c t + \theta)$  is to be reconstructed by multiplying  $v(t)$  by a signal derived from  $v^2(t)$ .

(a) Show that  $v^2(t)$  has a component at the frequency  $2f_c$ . Find its amplitude.

(b) If  $m(t)$  is bandlimited to  $f_M$  and has a probability density

$$f(m) = \frac{1}{\sqrt{2\pi}} e^{-m^2/2} \quad -\infty \leq m \leq \infty$$

find the expected value of the amplitude of the component of  $v^2(t)$  at  $2f_c$ .

3.4-1. The envelope detector shown in Fig. 3.4-2a is used to recover the signal  $m(t)$  from the AM signal  $v(t) = [1 + m(t)] \cos \omega_c t$ , where  $m(t)$  is a square wave taking on the values 0 and  $-0.5$  volt and having a period  $T \gg 1/f_c$ . Sketch the recovered signal if  $RC = T/20$  and  $4T$ .



3.4-2. (a) The waveform  $v(t) = (1 + m \cos \omega_m t) \cos \omega_c t$ , with  $m$  a constant ( $m \leq 1$ ), is applied to the diode demodulator of Fig. 3.4-2a. Show that, if the demodulator output is to follow the envelope of  $v(t)$ , it is required that at any time  $t_0$ :

$$\frac{1}{RC} \geq \omega_m \left( \frac{m \sin \omega_m t_0}{1 + m \cos \omega_m t_0} \right)$$

(b) Using the result of part (a), show that if the demodulator is to follow the envelope at all times then  $m$  must be less than or equal to the value of  $m_0$ , determined from the equation

$$RC = \frac{1}{\omega_m} \frac{\sqrt{1 - m_0^2}}{m_0}$$

(c) Draw, qualitatively, the form of the demodulator output when the condition specified in part (b) is not satisfied.

3.5-1. The signal  $v(t) = (1 + m \cos \omega_m t) \cos \omega_c t$  is detected using a diode envelope detector. Sketch the detector output when  $m = 2$ .

3.6-1. The signal  $v(t) = [1 + 0.2 \cos(\omega_M/3)t] \cos \omega_c t$  is demodulated using a square-law demodulator having the characteristic  $v_o = (v + 2)^2$ . The output  $v_o(t)$  is then filtered by an ideal low-pass filter having a cutoff frequency at  $f_M$  Hz. Sketch the amplitude-frequency characteristics of the output waveform in the frequency range  $0 \leq f \leq f_M$ .

3.6-2. Repeat Prob. 3.6-1 if the square-law demodulator is centered at the origin so that  $v_o = v^2$ .

3.6-3. The signal  $v(t) = [1 + m(t)] \cos \omega_c t$  is square-law detected by a detector having the characteristic  $v_o = v^2$ . If the Fourier transform of  $m(t)$  is a constant  $M_0$  extending from  $-f_M$  to  $+f_M$ , sketch the Fourier transform of  $v_o(t)$  in the frequency range  $-f_M < f < f_M$ . Hint: Convolution in the frequency domain is needed to find the Fourier transform of  $m^2(t)$ . See Prob. 1.12-2.

3.6-4. The signal  $v(t) = (1 + 0.1 \cos \omega_1 t + 0.1 \cos 2\omega_1 t) \cos \omega_c t$  is detected by a square-law detector,  $v_o = 2v^2$ . Plot the amplitude-frequency characteristic of  $v_o(t)$ .

3.9-1. (a) Show that the signal

$$v(t) = \sum_{i=1}^N [\cos \omega_c t \cos(\omega_i t + \theta_i) - \sin \omega_c t \sin(\omega_i t + \theta_i)]$$

is an SSB-SC signal ( $\omega_c \gg \omega_N$ ). Is it the upper or lower sideband?

(b) Write an expression for the missing sideband.

(c) Obtain an expression for the total DSB-SC signal.

3.9-2. The SSB signal in Prob. 3.9-1 is multiplied by  $\cos \omega_c t$  and then low-pass filtered to recover the modulation.

(a) Show that the modulation is completely recovered if the cutoff frequency of the low-pass filter  $f_0$  is  $f_M < f_0 < 2f_c$ .

(b) If the multiplying signal were  $\cos(\omega_c t + \theta)$ , find the recovered signal.

(c) If the multiplying signal were  $\cos(\omega_c + \Delta\omega)t$ , find the recovered signal. Assume that  $\Delta\omega \ll \omega_c$ .

3.9-3. Show that the squaring circuit shown in Fig. 3.3-1 will not permit the generation of a local oscillator signal capable of demodulating an SSB-SC signal.

3.10-1. A baseband signal, bandlimited to the frequency range 300 to 3000 Hz, is to be superimposed on a carrier of frequency of 40 MHz as a single-sideband modulation using the filter method. Assume that bandpass filters are available which will provide 40 dB of attenuation in a frequency interval which is about 1 percent of the filter center frequency. Draw a block diagram of a suitable system. At each point in the system draw plots indicating the spectral range occupied by the signal present there.

3.10-2. The system shown in Fig. 3.10-2 is used to generate a single-sideband signal. However, an ideal  $90^\circ$  phase-shifting network which is independent of a frequency is unattainable. The  $90^\circ$  phase shift is approximated by a lattice network having the transfer function

$$H(f) = e^{-j \arctan(f/30)}$$

The input to this network is  $m(t)$ , given by Eq. (3.10-3). If  $f_1 = 300$  Hz and  $f_M = 3000$  Hz show that  $H(f) \cong e^{-j\pi/2} e^{j30/f}$ , for  $f_1 \leq f \leq f_M$ .

3.10-3. In the SSB generating system of Fig. 3.10-2, the carrier phase-shift network produces a phase shift which differs from  $90^\circ$  by a small angle  $\alpha$ . Calculate the output waveform and point out the respects in which the output no longer meets the requirements for an SSB waveform. Assume that the input is a single spectral component  $\cos \omega_m t$ .

3.10-4. Repeat Prob. 3.10-3 except, assume instead, that the baseband phase-shift network produces a phase shift differing from  $90^\circ$  by a small angle  $\alpha$ .

3.11-1. A received SSB signal in which the modulation is a single spectral component has a normalized power of 0.5 volt<sup>2</sup>. A carrier is added to the signal, and the carrier plus signal are applied to a diode demodulator. The carrier amplitude is to be adjusted so that at the demodulator output 90 percent of the normalized power is in the recovered modulating waveform. Neglect dc components. Find the carrier amplitude required.

---

CHAPTER  
**FOUR**

---

## FREQUENCY-MODULATION SYSTEMS

In the amplitude-modulation systems described in Chap. 3, the modulator output consisted of a carrier which displayed variations in its amplitude. In the present chapter we discuss modulation systems in which the modulator output is of constant amplitude and in which the signal information is superimposed on the carrier through variations of the carrier frequency.

### 4.1 ANGLE MODULATION<sup>1</sup>

All the modulation schemes considered up to the present point have two principal features in common. In the first place, each spectral component of the baseband signal gives rise to one or two spectral components in the modulated signal. These components are separated from the carrier by a frequency difference equal to the frequency of the baseband component. Most importantly, the nature of the modulators is such that the spectral components which they produce depend only on the carrier frequency and the baseband frequencies. The amplitudes of the spectral components of the modulator output may depend on the amplitude of the input signals; however, the frequencies of the spectral components do not. In the second place, all the operations performed on the signal (addition, subtraction, and multiplication) are linear operations so that superposition applies. Thus, if a baseband signal  $m_1(t)$  introduces one spectrum of components into the modulated signal and a second signal  $m_2(t)$  introduces a second spectrum, the application of the sum  $m_1(t) + m_2(t)$  will introduce a spectrum which is the sum of the spectra separately introduced. All these systems are referred to under the designation "amplitude or linear modulation." This terminology must be taken

with some reservation, for we have noted that, at least in the special case of single sideband using modulation with a single sinusoid, there is no amplitude variation at all. And even more generally, when the amplitude of the modulated signal does vary, the carrier envelope need not have the waveform of the baseband signal.

We now turn our attention to a new type of modulation which is not characterized by the features referred to above. The spectral components in the modulated waveform depend on the amplitude as well as the frequency of the spectral components in the baseband signal. Furthermore, the modulation system is *not* linear and superposition does *not* apply. Such a system results when, in connection with a carrier of constant amplitude, the phase angle is made to respond in some way to a baseband signal. Such a signal has the form

$$v(t) = A \cos [\omega_c t + \phi(t)] \quad (4.1-1)$$

in which  $A$  and  $\omega_c$  are constant but in which the phase angle  $\phi(t)$  is a function of the baseband signal. Modulation of this type is called *angle modulation* for obvious reasons. It is also referred to as *phase modulation* since  $\phi(t)$  is the phase angle of the argument of the cosine function. Still another designation is *frequency modulation* for reasons to be discussed in the next section.

### 4.2 PHASE AND FREQUENCY MODULATION

To review some elementary ideas in connection with sinusoidal waveforms, let us recall that the function  $A \cos \omega_c t$  can be written as

$$A \cos \omega_c t = \text{real part } (Ae^{j\omega_c t}) \quad (4.2-1)$$

The function  $Ae^{j\theta}$  is represented in the complex plane by a phasor of length  $A$  and an angle  $\theta$  measured counterclockwise from the real axis. If  $\theta = \omega_c t$ , then the phasor rotates in the counterclockwise direction with an angular velocity  $\omega_c$ . With respect to a coordinate system which also rotates in the counterclockwise direction with angular velocity  $\omega_c$ , the phasor will be stationary. If in Eq. (4.1-1)  $\phi$  is actually not time-dependent but is a constant, then  $v(t)$  is to be represented precisely in the manner just described. But suppose  $\phi = \phi(t)$  does change with time and makes positive and negative excursions. Then  $v(t)$  would be represented by a phasor of amplitude  $A$  which runs ahead of and falls behind the phasor representing  $A \cos \omega_c t$ . We may, therefore, consider that the angle  $\omega_c t + \phi(t)$ , of  $v(t)$ , undergoes a *modulation* around the angle  $\theta = \omega_c t$ . The waveform of  $v(t)$  is, therefore, a representation of a signal which is *modulated in phase*.

If the phasor of angle  $\theta + \phi(t) = \omega_c t + \phi(t)$  alternately runs ahead of and falls behind the phasor  $\theta = \omega_c t$ , then the first phasor must alternately be rotating more, or less, rapidly than the second phasor. Therefore we may consider that the angular velocity of the phasor of  $v(t)$  undergoes a modulation around the nominal angular velocity  $\omega_c$ . The signal  $v(t)$  is, therefore, an angular-velocity-modulated waveform. The angular velocity associated with the argument of a sinusoidal function is equal to the time rate of change of the argument (i.e., the

angle) of the function. Thus we have that the instantaneous radial frequency  $\omega = d(\theta + \phi)/dt$ , and the corresponding frequency  $f = \omega/2\pi$  is

$$f = \frac{1}{2\pi} \frac{d}{dt} [\omega_c t + \phi(t)] = \frac{\omega_c}{2\pi} + \frac{1}{2\pi} \frac{d}{dt} \phi(t) \quad (4.2-2)$$

The waveform  $v(t)$  is, therefore, *modulated in frequency*.

In initial discussions of the sinusoidal waveform it is customary to consider such a waveform as having a fixed frequency and phase. In the present discussion we have generalized these concepts somewhat. To acknowledge this generalization, it is not uncommon to refer to the frequency  $f$  in Eq. (4.2-2) as the *instantaneous frequency* and  $\phi(t)$  as the *instantaneous phase*. If the frequency variation about the nominal frequency  $\omega_c$  is small, that is, if  $d\phi(t)/dt \ll \omega_c$ , then the resultant waveform will have an appearance which is readily recognizable as a "sine wave," albeit with a period which changes somewhat from cycle to cycle. Such a waveform is represented in Fig. 4.2-1. In this figure the modulating signal is a square wave. The frequency-modulated signal changes frequency whenever the modulation changes level.

Among the possibilities which suggest themselves for the design of a modulator are the following. We might arrange that the phase  $\phi(t)$  in Eq. (4.1-1) be directly proportional to the modulating signal, or we might arrange a direct proportionality between the modulating signal and the derivative,  $d\phi(t)/dt$ . From Eq. (4.2-2), with  $f_c = \omega_c/2\pi$

$$\frac{d\phi(t)}{dt} = 2\pi(f - f_c) \quad (4.2-3)$$

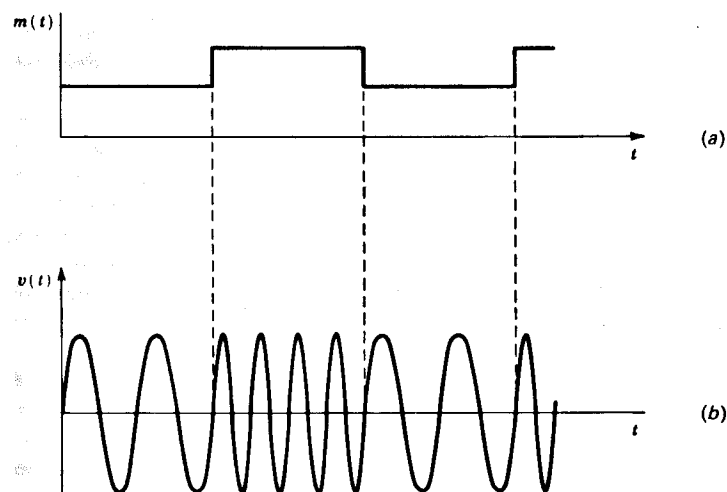


Figure 4.2-1 An angle-modulated waveform. (a) Modulating signal. (b) Frequently-modulated sinusoidal carrier signal.

where  $f$  is the instantaneous frequency. Hence in this latter case the proportionality is between modulating signal and the departure of the instantaneous frequency from the carrier frequency. Using standard terminology, we refer to the modulation of the first type as *phase modulation*, and the term *frequency modulation* refers only to the second type. On the basis of these definitions it is, of course, not possible to determine which type of modulation is involved simply from a visual examination of the waveform or from an analytical expression for the waveform. We would also have to be given the waveform of the modulating signal. This information is, however, provided in any practical communication system.

### 4.3 RELATIONSHIP BETWEEN PHASE AND FREQUENCY MODULATION

The relationship between phase and frequency modulation may be visualized further by a consideration of the diagrams of Fig. 4.3-1. In Fig. 4.3-1a the phase-modulator block represents a device which furnishes an output  $v(t)$  which is a carrier, phase-modulated by the input signal  $m_i(t)$ . Thus

$$v(t) = A \cos [\omega_c t + k'm_i(t)] \quad (4.3-1)$$

$k'$  being a constant. Let the waveform  $m_i(t)$  be derived as the integral of the modulating signal  $m(t)$  so that

$$m_i(t) = k'' \int_{-\infty}^t m(t) dt \quad (4.3-2)$$

in which  $k''$  is also a constant. Then with  $k = k'k''$  we have

$$v(t) = A \cos [\omega_c t + k \int_{-\infty}^t m(t) dt] \quad (4.3-3)$$

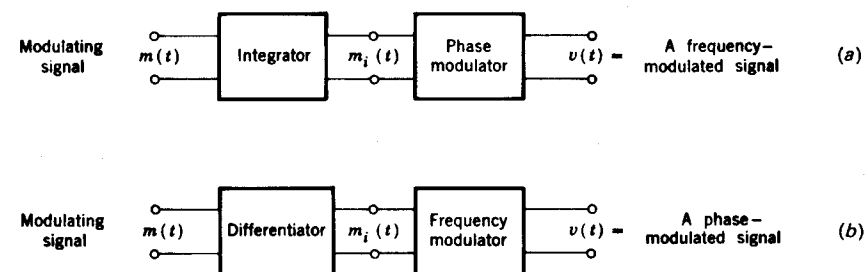


Figure 4.3-1 Illustrating the relationship between phase and frequency modulation.

The instantaneous angular frequency is

$$\omega = \frac{d}{dt} \left[ \omega_c t + k \int_{-\infty}^t m(t) dt \right] = \omega_c + km(t) \quad (4.3-4)$$

The deviation of the instantaneous frequency from the carrier frequency  $\omega_c/2\pi$  is

$$\nu \equiv f - f_c = \frac{k}{2\pi} m(t) \quad (4.3-5)$$

Since the deviation of the instantaneous frequency is directly proportional to the modulating signal, the combination of *integrator* and *phase modulator* of Fig. 4.3-1a constitutes a device for producing a *frequency-modulated* output. Similarly, the combination in Fig. 4.3-1b of the differentiator and frequency modulator generates a *phase-modulated* output, i.e., a signal whose phase departure from the carrier is proportional to the modulating signal.

In summary, we have referred generally to the waveform given by Eq. (4.1-1) as an *angle-modulated* waveform, an appropriate designation when we have no interest in, or information about, the modulating signal. When  $\phi(t)$  is proportional to the modulating signal  $m(t)$ , we use the designation *phase modulation* or PM. When the time derivative of  $\phi(t)$  is proportional to  $m(t)$ , we use the term *frequency modulation* or FM. In an FM waveform, the form of Eq. (4.3-3) is of special interest, since here the instantaneous frequency deviation is directly proportional to the signal  $m(t)$  which appears explicitly in the expression. In general usage, however, we find that such precision of language is not common. Very frequently the terms angle modulation, phase modulation, and frequency modulation are used rather interchangeably and without reference to, or even interest in, the modulating signal.

#### 4.4 PHASE AND FREQUENCY DEVIATION

In the waveform of Eq. (4.1-1) the maximum value attained by  $\phi(t)$ , that is, the maximum phase deviation of the total angle from the carrier angle  $\omega_c t$ , is called the *phase deviation*. Similarly the maximum departure of the instantaneous frequency from the carrier frequency is called the *frequency deviation*.

When the angular (and consequently the frequency) variation is sinusoidal with frequency  $f_m$ , we have, with  $\omega_m = 2\pi f_m$

$$v(t) = A \cos(\omega_c t + \beta \sin \omega_m t) \quad (4.4-1)$$

where  $\beta$  is the peak amplitude of  $\phi(t)$ . In this case  $\beta$  which is the maximum phase deviation, is usually referred to as the *modulation index*. The instantaneous frequency is

$$f = \frac{\omega_c}{2\pi} + \frac{\beta \omega_m}{2\pi} \cos \omega_m t \quad (4.4-2a)$$

$$= f_c + \beta f_m \cos \omega_m t \quad (4.4-2b)$$

The maximum frequency deviation is defined as  $\Delta f$  and is given by

$$\Delta f = \beta f_m \quad (4.4-3)$$

Equation (4.4-1) can, therefore, be written

$$v(t) = A \cos \left( \omega_c t + \frac{\Delta f}{f_m} \sin \omega_m t \right) \quad (4.4-4)$$

While the instantaneous frequency  $f$  lies in the range  $f_c \pm \Delta f$ , it should not be concluded that all spectral components of such a signal lie in this range. We consider next the spectral pattern of such an angle-modulated waveform.

#### 4.5 SPECTRUM OF AN FM SIGNAL: SINUSOIDAL MODULATION

In this section we shall look into the frequency spectrum of the signal

$$v(t) = \cos(\omega_c t + \beta \sin \omega_m t) \quad (4.5-1)$$

which is the signal of Eq. (4.4-1) with the amplitude arbitrarily set at unity as a matter of convenience. We have

$$\begin{aligned} \cos(\omega_c t + \beta \sin \omega_m t) &= \cos \omega_c t \cos(\beta \sin \omega_m t) \\ &\quad - \sin \omega_c t \sin(\beta \sin \omega_m t) \end{aligned} \quad (4.5-2)$$

Consider now the expression  $\cos(\beta \sin \omega_m t)$  which appears as a factor on the right-hand side of Eq. (4.5-2). It is an *even*, periodic function having an angular frequency  $\omega_m$ . Therefore it is possible to expand this expression in a Fourier series in which  $\omega_m/2\pi$  is the fundamental frequency. We shall not undertake the evaluation of the coefficients in the Fourier expansion of  $\cos(\beta \sin \omega_m t)$  but shall instead simply write out the results. The coefficients are, of course, functions of  $\beta$ , and, since the function is *even*, the coefficients of the odd harmonics are zero. The result is

$$\begin{aligned} \cos(\beta \sin \omega_m t) &= J_0(\beta) + 2J_2(\beta) \cos 2\omega_m t + 2J_4(\beta) \cos 4\omega_m t \\ &\quad + \cdots + 2J_{2n}(\beta) \cos 2n\omega_m t + \cdots \end{aligned} \quad (4.5-3)$$

while for  $\sin(\beta \sin \omega_m t)$ , which is an *odd* function, we find the expansion contains only odd harmonics and is given by

$$\begin{aligned} \sin(\beta \sin \omega_m t) &= 2J_1(\beta) \sin \omega_m t + 2J_3(\beta) \sin 3\omega_m t \\ &\quad + \cdots + 2J_{2n-1}(\beta) \sin(2n-1)\omega_m t + \cdots \end{aligned} \quad (4.5-4)$$

The functions  $J_n(\beta)$  occur often in the solution of engineering problems. They are known as Bessel functions of the first kind and of order  $n$ . The numerical values of  $J_n(\beta)$  are tabulated in texts of mathematical tables.<sup>2</sup>

Putting the results given in Eqs. (4.5-3) and (4.5-4) back into Eq. (4.5-2) and using the identities

$$\cos A \cos B = \frac{1}{2} \cos (A - B) + \frac{1}{2} \cos (A + B) \quad (4.5-5)$$

$$\sin A \sin B = \frac{1}{2} \cos (A - B) - \frac{1}{2} \cos (A + B) \quad (4.5-6)$$

we find that  $v(t)$  in Eq. (4.5-1) becomes

$$\begin{aligned} v(t) = & J_0(\beta) \cos \omega_c t - J_1(\beta) [\cos (\omega_c - \omega_m)t - \cos (\omega_c + \omega_m)t] \\ & + J_2(\beta) [\cos (\omega_c - 2\omega_m)t + \cos (\omega_c + 2\omega_m)t] \\ & - J_3(\beta) [\cos (\omega_c - 3\omega_m)t - \cos (\omega_c + 3\omega_m)t] \\ & + \dots \end{aligned} \quad (4.5-7)$$

Observe that the spectrum is composed of a carrier with an amplitude  $J_0(\beta)$  and a set of sidebands spaced symmetrically on either side of the carrier at frequency separations of  $\omega_m$ ,  $2\omega_m$ ,  $3\omega_m$ , etc. In this respect the result is unlike that which prevails in the amplitude-modulation systems discussed earlier, since in AM a sinusoidal modulating signal gives rise to only one sideband or one pair of sidebands. A second difference, which is left for verification by the student (Prob. 4.5-1), is that the present modulation system is nonlinear, as anticipated from the discussion of Sec. 4.1.

## 4.6 SOME FEATURES OF THE BESSEL COEFFICIENTS

Several of the Bessel functions which determine the amplitudes of the spectral components in the Fourier expansion are plotted in Fig. 4.6-1. We note that, at  $\beta = 0$ ,  $J_0(0) = 1$ , while all other  $J_n$ 's are zero. Thus, as expected when there is no modulation, only the carrier, of normalized amplitude unity, is present, while all sidebands have zero amplitude. When  $\beta$  departs slightly from zero,  $J_1(\beta)$  acquires a magnitude which is significant in comparison with unity, while all higher-order  $J$ 's are negligible in comparison. That such is the case may be seen either from Fig. 4.6-1 or from the approximations<sup>3</sup> which apply when  $\beta \ll 1$ , that is,

$$J_0(\beta) \cong 1 - \left(\frac{\beta}{2}\right)^2 \quad (4.6-1)$$

$$J_n(\beta) \cong \frac{1}{n!} \left(\frac{\beta}{2}\right)^n \quad n \neq 0 \quad (4.6-2)$$

Accordingly, for  $\beta$  very small, the FM signal is composed of a carrier and a single pair of sidebands with frequencies  $\omega_c \pm \omega_m$ . An FM signal which is so constituted, that is, a signal where  $\beta$  is small enough so that only a single sideband pair is of significant magnitude, is called a *narrowband* FM signal. We see further, in Fig. 4.6-1, as  $\beta$  becomes somewhat larger, that the amplitude  $J_1$  of the first sideband pair increases and that also the amplitude  $J_2$  of the second sideband pair

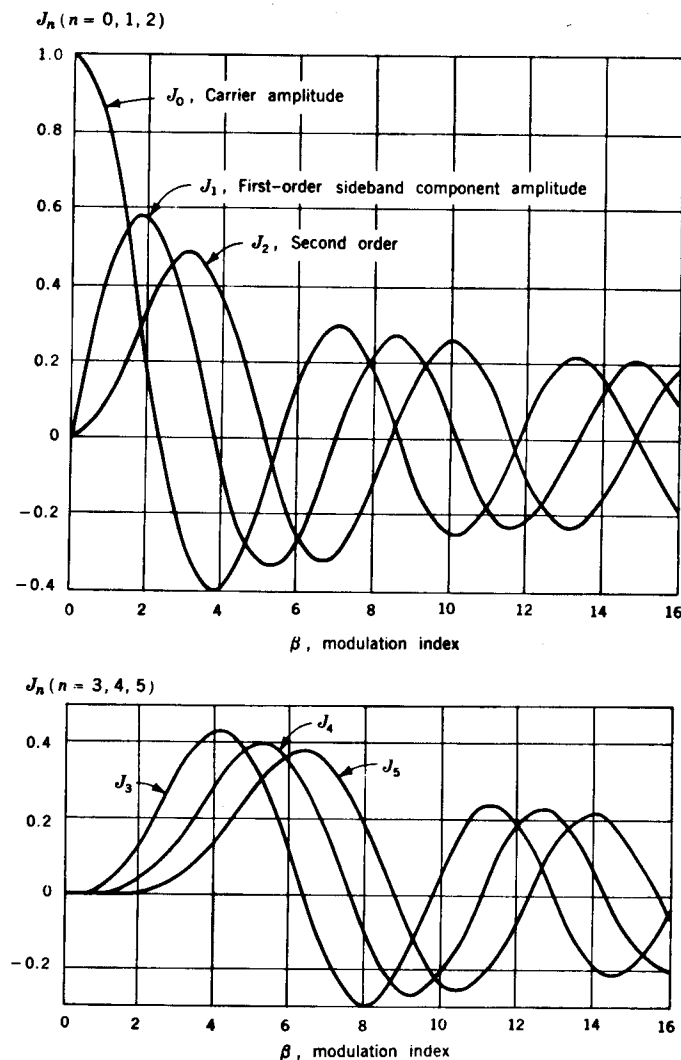


Figure 4.6-1 The Bessel functions  $J_n(\beta)$  plotted as a function of  $\beta$  for  $n = 0, 1, 2, \dots, 5$ .

becomes significant. Further, as  $\beta$  continues to increase,  $J_3$ ,  $J_4$ , etc. begin to acquire significant magnitude, giving rise to sideband pairs at frequencies  $\omega_c \pm 2\omega_m$ ,  $\omega_c \pm 3\omega_m$ , etc.

Another respect in which FM is unlike the linear-modulation schemes described earlier is that in an FM signal the amplitude of the spectral component at the carrier frequency is not constant independent of  $\beta$ . It is to be expected that such should be the case on the basis of the following considerations. The envelope of an FM signal has a constant amplitude. Therefore the power of such a signal is a constant independent of the modulation, since the power of a periodic waveform depends only on the square of its amplitude and not on its frequency. The power of a unit amplitude signal, as in Eq. (4.5-1), is  $P_v = \frac{1}{2}$  and is independent of  $\beta$ . When the carrier is modulated to generate an FM signal, the power in the sidebands may appear only at the expense of the power originally in the carrier. Another way of arriving at the same conclusion is to make use of the identity<sup>3</sup>  $J_0^2 + 2J_1^2 + 2J_2^2 + 2J_3^2 + \dots = 1$ . We calculate the power  $P_v$  by squaring  $v(t)$  in Eq. (4.5-7) and then averaging  $v^2(t)$ . Keeping in mind that cross-product terms average to zero, we find, independently of  $\beta$ , that

$$P_v = \frac{1}{2} \left( J_0^2 + 2 \sum_{n=1}^{\infty} J_n^2 \right) = \frac{1}{2} \quad (4.6-3)$$

as expected. We observe in Fig. 4.6-1 that, at various values of  $\beta$ ,  $J_0(\beta) = 0$ . At these values of  $\beta$  all the power is in the sidebands and none in the carrier.

## 4.7 BANDWIDTH OF A SINUSOIDALLY MODULATED FM SIGNAL

In principle, when an FM signal is modulated, the number of sidebands is infinite and the bandwidth required to encompass such a signal is similarly infinite in extent. As a matter of practice, it turns out that for any  $\beta$ , so large a fraction of the total power is confined to the sidebands which lie within some finite bandwidth that no serious distortion of the signal results if the sidebands outside this bandwidth are lost. We see in Fig. 4.6-1 that, except for  $J_0(\beta)$ , each  $J_n(\beta)$  hugs the zero axis initially and that as  $n$  increases, the corresponding  $J_n$  remains very close to the zero axis up to a larger value of  $\beta$ . For any value of  $\beta$  only those  $J_n$  need be considered which have succeeded in making a significant departure from the zero axis. How many such sideband components need to be considered may be seen from an examination of Table 4.7-1 where  $J_n(\beta)$  is tabulated for various values of  $n$  and of  $\beta$ .

It is found experimentally that the distortion resulting from bandlimiting an FM signal is tolerable as long as 98 percent or more of the power is passed by the bandlimiting filter. This definition of the bandwidth of a filter is, admittedly, somewhat vague, especially since the term "tolerable" means different things in different applications. However, using this definition for bandwidth, one can proceed with an initial tentative design of a system. When the system is built, the

Table 4.7-1 Values of the Bessel functions  $J_n(\beta)$  for various orders  $n$  and integral values of  $\beta$

| $\beta \backslash n$ | 0        | 1        | 2        | 3        | 4        | 5        | 6        | 7         | 8        | 9        | 10       |
|----------------------|----------|----------|----------|----------|----------|----------|----------|-----------|----------|----------|----------|
| 0                    | 0.7652   | 0.2239   | 0.2239   | -0.2601  | -0.3971  | -0.1776  | 0.1506   | 0.3001    | 0.1717   | -0.09033 | -0.2459  |
| 1                    | 0.4401   | 0.5767   | 0.5767   | 0.3391   | -0.06604 | -0.3276  | -0.2767  | -0.004683 | 0.2346   | 0.2453   | 0.04347  |
| 2                    | 0.1149   | 0.3528   | 0.3528   | 0.4861   | 0.3641   | 0.04657  | -0.2429  | -0.3014   | -0.1130  | 0.1448   | 0.2546   |
| 3                    | 0.01956  | 0.1289   | 0.1289   | 0.3091   | 0.4302   | 0.3648   | 0.1148   | -0.1676   | -0.2911  | -0.1809  | 0.03838  |
| 4                    | 0.002477 | 0.03400  | 0.03400  | 0.1320   | 0.2811   | 0.3912   | 0.3576   | 0.1578    | -0.1054  | -0.2655  | -0.2196  |
| 5                    |          | 0.007040 | 0.007040 | 0.04303  | 0.1321   | 0.2611   | 0.3621   | 0.3479    | 0.1858   | -0.05504 | -0.2341  |
| 6                    |          | 0.001202 | 0.001202 | 0.01139  | 0.04909  | 0.1310   | 0.2458   | 0.3392    | 0.3376   | 0.2043   | -0.01446 |
| 7                    |          |          |          | 0.002547 | 0.01518  | 0.05338  | 0.1296   | 0.2336    | 0.3206   | 0.3275   | 0.2167   |
| 8                    |          |          |          |          | 0.004029 | 0.01841  | 0.05653  | 0.1280    | 0.2235   | 0.3051   | 0.3179   |
| 9                    |          |          |          |          |          | 0.005520 | 0.02117  | 0.05892   | 0.1263   | 0.2149   | 0.2919   |
| 10                   |          |          |          |          |          | 0.001468 | 0.006964 | 0.02354   | 0.06077  | 0.1247   | 0.2075   |
| 11                   |          |          |          |          |          |          | 0.002048 | 0.008335  | 0.02560  | 0.06222  | 0.1231   |
| 12                   |          |          |          |          |          |          |          | 0.002656  | 0.009624 | 0.02739  | 0.06337  |
| 13                   |          |          |          |          |          |          |          |           | 0.003275 | 0.01083  | 0.02897  |
| 14                   |          |          |          |          |          |          |          |           |          | 0.003895 | 0.01196  |
| 15                   |          |          |          |          |          |          |          |           |          | 0.001286 | 0.004508 |
| 16                   |          |          |          |          |          |          |          |           |          |          | 0.001567 |

bandwidth may thereafter be readjusted, if necessary. In each column of Table 4.7-1, a line has been drawn after the entries which account for at least 98 percent of the power. To illustrate this point, consider  $\beta = 1$ . Then the power contained in the terms  $n = 0, 1$ , and  $2$  is

$$P = \frac{1}{2}J_0^2(1) + J_1^2(1) + J_2^2(1) \\ = 0.289 + 0.193 + 0.013 = 0.495 \quad (4.7-1)$$

The sum 0.495 is 99 percent of the power in the FM signal, which is  $\frac{1}{2}$ .

We note that the horizontal lines in Table 4.7-1, which indicate the value of  $n$  for 98 percent power transmission, always occur just after  $n = \beta + 1$ . Thus, for sinusoidal modulation the bandwidth required to transmit or receive the FM signal is

$$B = 2(\beta + 1)f_m \quad (4.7-2)$$

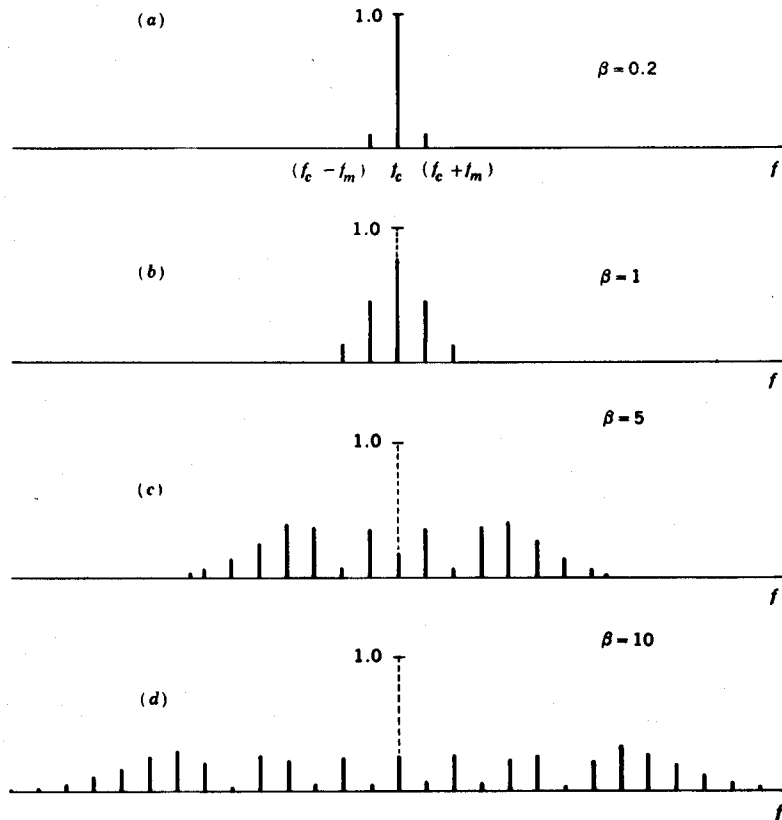


Figure 4.7-1 The spectra of sinusoidally modulated FM signals for various values of  $\beta$ .

By way of example, when  $\beta = 5$ , the sideband components furthest from the carrier which have adequate amplitude to require consideration are those which occur at frequencies  $f_c \pm 6f_m$ . From the table of Bessel functions published in Jahnke and Emde<sup>2</sup> it may be verified on a numerical basis that the rule given in Eq. (4.7-2) holds without exception up to  $\beta = 29$ , which is the largest value of  $\beta$  for which  $J_m$  is tabulated there.

Using Eq. (4.4-3), we may put Eq. (4.7-2) in a form which is more immediately significant. We find

$$B = 2(\Delta f + f_m) \quad (4.7-3)$$

Expressed in words, *the bandwidth is twice the sum of the maximum frequency deviation and the modulating frequency*. This rule for bandwidth is called *Carson's rule*.

We deduced Eqs. (4.7-2) and (4.7-3) as a generalization from Table 4.7-1, which begins with  $\beta = 1$ . We may note, however, that the bandwidth approximation applies quite well even when  $\beta \ll 1$ . For, in that case, we find that Eq. (4.7-2) gives  $B = 2f_m$ , which we know to be correct from our earlier discussion of narrowband FM.

The spectra of several FM signals with sinusoidal modulation are shown in Fig. 4.7-1 for various values of  $\beta$ . These spectra are constructed directly from the entries in Table 4.7-1 except that the signs of the terms have been ignored. The spectral lines have, in every case, been drawn upward even when the corresponding entry is negative. Hence, the lines represent the *magnitudes* only of the spectral components. Not all spectral components have been drawn. Those, far removed from the carrier, which are too small to be drawn conveniently to scale, have been omitted.

#### 4.8 EFFECT OF THE MODULATION INDEX $\beta$ ON BANDWIDTH

The modulation index  $\beta$  plays a role in FM which is not unlike the role played by the parameter  $m$  in connection with AM. In the AM case, and for sinusoidal modulation, we established that to avoid distortion we must observe  $m = 1$  as an upper limit. It was also apparent that when it is feasible to do so, it is advantageous to adjust  $m$  to be close to unity, that is, 100 percent modulation; by so doing, we keep the magnitude of the recovered baseband signal at a maximum. On this same basis we expect the advantage to lie with keeping  $\beta$  as large as possible. For, again, the larger is  $\beta$ , the stronger will be the recovered signal. While in AM the constraint that  $m \leq 1$  is imposed by the necessity to avoid distortion, there is no similar absolute constraint on  $\beta$ .

There is, however, a constraint which needs to be imposed on  $\beta$  for a different reason. From Eq. (4.7-2) for  $\beta \gg 1$  we have  $B \approx 2\beta f_m$ . Therefore the maximum value we may allow for  $\beta$  is determined by the maximum allowable bandwidth and the modulation frequency. In comparing AM with FM, we may

then note, in review, that in AM the recovered modulating signal may be made progressively larger subject to the onset of distortion in a manner which keeps the occupied bandwidth constant. In FM there is no similar limit on the modulation, but increasing the magnitude of the recovered signal is achieved at the expense of bandwidth. A more complete comparison is deferred to Chaps. 8 and 9, where we shall take account of the presence of noise and also of the relative power required for transmission.

#### 4.9 SPECTRUM OF "CONSTANT BANDWIDTH" FM

Let us consider that we are dealing with a modulating signal voltage  $v_m \cos 2\pi f_m t$  with  $v_m$  the peak voltage. In a phase-modulating system the phase angle  $\phi(t)$  would be proportional to this modulating signal so that  $\phi(t) = k'v_m \cos 2\pi f_m t$ , with  $k'$  a constant. The phase deviation is  $\beta = k'v_m$ , and, for constant  $v_m$ , the bandwidth occupied increases linearly with modulating frequency since  $B \cong 2\beta f_m = 2k'v_m f_m$ . We may avoid this variability of bandwidth with modulating frequency by arranging that  $\phi(t) = (k/2\pi f_m)v_m \sin 2\pi f_m t$  ( $k$  a constant). For, in this case

$$\beta = \frac{kv_m}{2\pi f_m} \quad (4.9-1)$$

and the bandwidth is  $B \cong (2k/2\pi)v_m$ , independently of  $f_m$ . In this latter case, however, the instantaneous frequency is  $\omega = \omega_c + kv_m \cos 2\pi f_m t$ . Since the instantaneous frequency is proportional to the modulating signal, the initially angle-modulated signal has become a frequency-modulated signal. Thus a signal intended to occupy a nominally constant bandwidth is a frequently-modulated rather than an angle-modulated signal.

In Fig. 4.9-1 we have drawn the spectrum for three values of  $\beta$  for the condition that  $\beta f_m$  is kept constant. The nominal bandwidth  $B \cong 2\Delta f = 2\beta f_m$  is consequently constant. The amplitude of the unmodulated carrier at  $f_c$  is shown by a dashed line. Note that the extent to which the actual bandwidth extends beyond the nominal bandwidth is greatest for small  $\beta$  and large  $f_m$  and is least for large  $\beta$  and small  $f_m$ .

In commercial FM broadcasting, the Federal Communications Commission allows a frequency deviation  $\Delta f = 75$  kHz. If we assume that the highest audio frequency to be transmitted is 15 kHz, then at this frequency  $\beta = \Delta f/f_m = 75/15 = 5$ . For all other modulation frequencies  $\beta$  is larger than 5. When  $\beta = 5$ , there are  $\beta + 1 = 6$  significant sideband pairs so that at  $f_m = 15$  kHz the bandwidth required is  $B = 2 \times 6 \times 15 = 180$  kHz, which is to be compared with  $2\Delta f = 150$  kHz. When  $\beta = 20$ , there are 21 significant sideband pairs, and  $B = 2 \times 21 \times 15/4 = 157.5$  kHz. In the limiting case of very large  $\beta$  and correspondingly very small  $f_m$ , the actual bandwidth becomes equal to the nominal bandwidth  $2\Delta f$ .

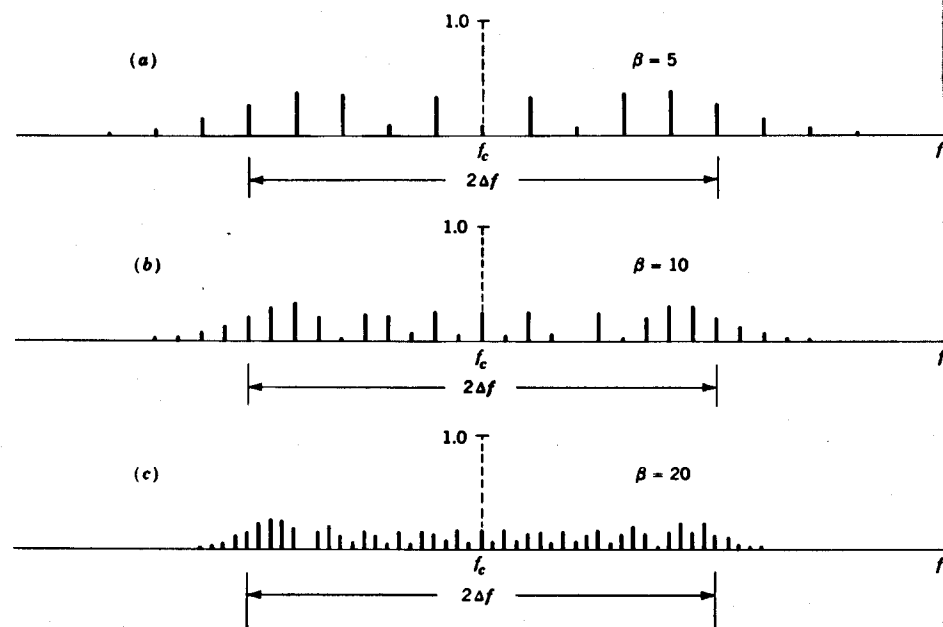


Figure 4.9-1 Spectra of sinusoidally modulated FM signals. The nominal bandwidth  $B \approx 2\beta f_m = 2\Delta f$  is kept fixed.

#### 4.10 PHASOR DIAGRAM FOR FM SIGNALS

With the aid of a phasor diagram we shall be able to arrive at a rather physically intuitive understanding of how so odd an assortment of sidebands as in Eq. (4.5-7) yields an FM signal of constant amplitude. The diagram will also make clear the difference between AM and narrowband FM (NBFM). In both of these cases there is only a single pair of sideband components.

Let us consider first the case of narrowband FM. From Eqs. (4.4-1), (4.6-1), and (4.6-2) we have for  $\beta \ll 1$  that

$$v(t) = \cos(\omega_c t + \beta \sin \omega_m t) \quad (4.10-1a)$$

$$\cong \cos \omega_c t - \frac{\beta}{2} \cos(\omega_c - \omega_m)t + \frac{\beta}{2} \cos(\omega_c + \omega_m)t \quad (4.10-1b)$$

Refer to Fig. 4.10-1a. Assuming a coordinate system which rotates counterclockwise at an angular velocity  $\omega_c$ , the phasor for the carrier-frequency term in Eq. (4.10-1) is fixed and oriented in the horizontal direction. In the same coordinate system, the phasor for the term  $(\beta/2) \cos(\omega_c + \omega_m)t$  rotates in a counterclockwise direction at an angular velocity  $\omega_m$ , while the phasor for the term



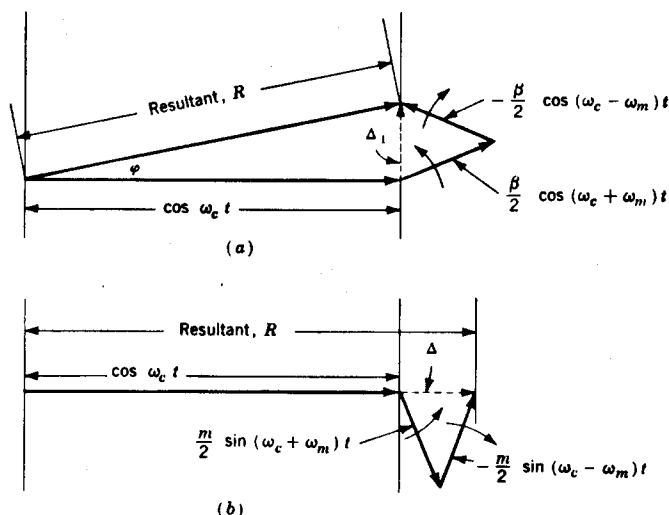


Figure 4.10-1 (a) Phasor diagram for a narrowband FM signal. (b) Phasor diagram for an AM signal.

$-(\beta/2) \cos(\omega_c - \omega_m)t$  rotates in a clockwise direction, also at the angular velocity  $\omega_m$ . At the time  $t = 0$ , both phasors, which represent the sideband components, have maximum projections in the horizontal direction. At this time one is parallel to, and one is antiparallel to, the phasor representing the carrier, so that the two cancel. The situation depicted in Fig. 4.10-1a corresponds to a time shortly after  $t = 0$ . At this time, the rotation of the sideband phasors which are in opposite directions, as indicated by the curved arrows, have given rise to a sum phasor  $\Delta_1$ . In the coordinate system in which the carrier phasor is stationary, the phasor  $\Delta_1$  always stands perpendicularly to the carrier phasor and has the magnitude

$$\Delta_1 = \beta \sin \omega_m t \quad (4.10-2)$$

The carrier, now slightly reduced in amplitude, and  $\Delta_1$  combine to give rise to a resultant  $R$ . The angular departure of  $R$  from the carrier phasor is  $\phi$ . It is readily seen from Fig. 4.10-1a that since  $\beta \ll 1$ , the maximum value of  $\phi \approx \tan \phi = \beta$ , as is to be expected. The small variation in the amplitude of the resultant which appears in Fig. 4.10-1a is only the result of the fact that we have neglected higher-order sidebands.

Now let us consider the phasor diagram for AM. The AM signal is

$$(1 + m \sin \omega_m t) \cos \omega_c t = \cos \omega_c t + \frac{m}{2} \sin(\omega_c + \omega_m)t - \frac{m}{2} \sin(\omega_c - \omega_m)t \quad (4.10-3)$$

and the individual terms are represented as phasors in Fig. 4.10-1b. Comparing Eqs. (4.10-1) and (4.10-3), we see that there is a  $90^\circ$  phase shift in the phases of the sidebands between the FM and AM cases. In Fig. 4.10-1b the sum  $\Delta$  of the sideband phasors is given by

$$\Delta = m \sin \omega_m t \quad (4.10-4)$$

The important difference between the FM and AM cases is that in the former the sum  $\Delta_1$  is always perpendicular to the carrier phasor, while in the latter the sum  $\Delta$  is always parallel to the carrier phasor. Hence in the AM case, the resultant  $R$  does not rotate with respect to the carrier phasor but instead varies in amplitude between  $1 + m$  and  $1 - m$ .

Another way of looking at the difference between AM and NBFM is to note that in NBFM where  $\beta \ll 1$

$$v(t) \approx \cos \omega_c t - \beta \sin \omega_m t \sin \omega_c t \quad (4.10-5)$$

while in AM

$$v(t) = \cos \omega_c t + m \sin \omega_m t \cos \omega_c t \quad (4.10-6)$$

Note that in NBFM the first term is  $\cos \omega_c t$ , while the second term involves  $\sin \omega_c t$ , a *quadrature* relationship. In AM both first and second terms involve  $\cos \omega_c t$ , an *in-phase* relationship.

To return now to the FM case and to Fig. 4.10-1a, the following point is worth noting. When the angle  $\phi$  completes a full cycle, that is,  $\Delta_1$  varies from  $+\beta$  to  $-\beta$  and back again to  $+\beta$ , the magnitude of the resultant  $R$  will have executed *two* full cycles. For  $R$  is a maximum at  $\Delta_1 = \beta$ , a minimum at  $\Delta_1 = 0$ , a maximum again when  $\Delta_1 = -\beta$ , and so on. On this basis, it may well be expected that if an additional sideband pair is to be added to the first to make  $R$  more nearly constant, this new pair must give rise to a resultant  $\Delta_2$  which varies at the frequency  $2\omega_m$ . Thus, we are not surprised to find that as the phase deviation  $\beta$  increases, a sideband pair comes into existence at the frequencies  $\omega_c \pm 2\omega_m$ .

As long as we depend on the first-order sideband pair only, we see from Fig. 4.10-1a that  $\phi$  cannot exceed  $90^\circ$ . A deviation of such magnitude is hardly adequate. For consider, as above, that  $\Delta f = 75$  kHz and that  $f_m = 50$  Hz. Then  $\omega_m = 75,000/50 = 1500$  rad, and the resultant  $R$  must, in this case, spin completely about  $1500/2\pi$  or about 240 times. Such wild whirling is made possible through the effect of the higher-order sidebands. As noted, the first-order sideband pair gives rise to a phasor  $\Delta_1 = J_1(\beta) \sin \omega_m t$ , which phasor is perpendicular to the carrier phasor. It may also be established by inspection of Eq. (4.5-7) that the second-order sideband pair gives rise to a phasor  $\Delta_2 = J_2(\beta) \cos 2\omega_m t$  and that this phasor is *parallel* to the carrier phasor. Continuing, we easily establish that all odd-numbered sideband pairs give rise to phasors

$$\Delta_n = J_n(\beta) \sin n\omega_m t \quad n \text{ odd} \quad (4.10-7)$$

which are perpendicular to the carrier phasor, while all even-numbered sideband pairs give rise to phasors

$$\Delta_n = J_n(\beta) \cos n\omega_m t \quad n \text{ even} \quad (4.10-8)$$

which are parallel to the carrier phasor. Thus, phasors  $\Delta_1, \Delta_2, \Delta_3$ , etc., alternately perpendicular and parallel to the carrier phasor, are added to carry the end point of the resultant phasor  $R$  completely around as many times as may be required, while maintaining  $R$  at constant magnitude. It is left as an exercise for the student to show by typical examples how the superposition of a carrier and sidebands may swing a constant-amplitude resultant around through an arbitrary angle (Prob. 4.10-2).

#### 4.11 SPECTRUM OF NARROWBAND ANGLE MODULATION: ARBITRARY MODULATION

Previously we considered the spectrum, in NBFM, which is produced by sinusoidal modulation. We found that, just as in AM, such modulation gives rise to two sidebands at frequencies  $\omega_c + \omega_m$  and  $\omega_c - \omega_m$ . We extend the result now to an arbitrary modulating waveform.

We may readily verify (Prob. 4.11-1) that superposition applies in narrowband angle modulation just as it does to AM. That is, if  $\beta_1 \sin \omega_1 t + \beta_2 \sin \omega_2 t$  is substituted in Eq. (4.10-1a) in place of  $\beta \sin \omega_m t$ , the sidebands which result are the sum of the sidebands that would be yielded by either modulation alone. Hence even if a modulating signal of waveform  $m(t)$ , with a continuous distribution of spectral components, is used in either AM or narrowband angle modulation the forms of the sideband spectra will be the same in the two cases.

More formally, we have in AM, when the modulating waveform is  $m(t)$ , the signal is

$$v_{AM}(t) = A[1 + m(t)] \cos \omega_c t = A \cos \omega_c t + Am(t) \cos \omega_c t \quad (4.11-1)$$

Let us assume, for simplicity, that  $m(t)$  is a finite energy waveform with a Fourier transform  $M(j\omega)$ . We use the theorem that if the Fourier transform  $\mathcal{F}[m(t)] = M(j\omega)$ , then  $\mathcal{F}[m(t) \cos \omega_c t]$  is as given in Eq. (3.2-4). We then find that

$$\mathcal{F}[v_{AM}(t)] = \frac{A}{2} [\delta(\omega + \omega_c) + \delta(\omega - \omega_c)] + \frac{A}{2} [M(j\omega + j\omega_c) + M(j\omega - j\omega_c)] \quad (4.11-2)$$

The narrowband angle-modulation signal of Eq. (4.10-1), except of amplitude  $A$  and with phase modulation  $m(t)$ , may be written, for  $|m(t)| \ll 1$ ,

$$v_{PM}(t) \cong A \cos \omega_c t - Am(t) \sin \omega_c t \quad (4.11-3)$$

so that

$$\mathcal{F}[v_{PM}(t)] = \frac{A}{2} [\delta(\omega + \omega_c) + \delta(\omega - \omega_c)] + \frac{A}{2} e^{-jn/2} [M(j\omega + j\omega_c) - M(j\omega - j\omega_c)] \quad (4.11-4)$$

Comparing Eq. (4.11-2) with Eq. (4.11-4), we observe that

$$|\mathcal{F}[v_{AM}]|^2 = |\mathcal{F}[v_{PM}]|^2 \quad (4.11-5)$$

Thus, if we were to make plots of the energy spectral densities of  $v_{AM}(t)$  and of  $v_{PM}(t)$ , we would find them identical. Similarly, if  $m(t)$  were a signal of finite power, we would find that plots of power spectral density would be the same.

#### 4.12 SPECTRUM OF WIDEBAND FM (WBFM): ARBITRARY MODULATION<sup>4</sup>

In this section we engage in a heuristic discussion of the spectrum of a wideband FM signal. We shall not be able to deduce the spectrum with the precision that is possible in the NBFM case described in the previous section. As a matter of fact, we shall be able to do no more than to deduce a means of expressing approximately the power spectral density of a WBFM signal. But this result is important and useful.

Previously, to characterize an FM signal as being narrowband or wideband, we had used the parameter  $\beta \equiv \Delta f/f_m$ , where  $\Delta f$  is the frequency deviation and  $f_m$  the frequency of the sinusoidal modulating signal. The signal was then NBFM or WBFM depending on whether  $\beta \ll 1$  or  $\beta \gg 1$ . Alternatively we distinguished one from the other on the basis of whether one or very many sidebands were produced by each spectral component of the modulating signal, and on the basis of whether or not superposition applies. We consider now still another alternative.

Let the symbol  $\nu \equiv f - f_c$  represent the frequency difference between the instantaneous frequency  $f$  and the carrier frequency  $f_c$ ; that is,  $\nu(t) = (k/2\pi)m(t)$  [see Eq. (4.3-5)]. The period corresponding to  $\nu$  is  $T = 1/\nu$ . As  $f$  varies, so also will  $\nu$  and  $T$ . The frequency  $\nu$  is the frequency with which the resultant phasor  $R$  in Fig. 4.10-1 rotates in the coordinate system in which the carrier phasor is fixed. In WBFM this resultant phasor rotates through many complete revolutions, and its speed of rotation does not change radically from revolution to revolution. Since, the resultant  $R$  is constant, then if we were to examine the plot as a function of time of the projection of  $R$  in, say, the horizontal direction, we would recognize it as a sinusoidal waveform because its frequency would be changing very slowly. No appreciable change in frequency would take place during the course of a cycle. Even a long succession of cycles would give the appearance of being of rather constant frequency. In NBFM, on the other hand, the phasor  $R$  simply oscillates about the position of the carrier phasor. Even though, in this case, we may still formally calculate a frequency  $\nu$ , there is no corresponding time interval